

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/107285>

**Copyright and reuse:**

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

**Isolation and expression of a plastid  $\alpha$  chaperonin  
cDNA sequence from *Triticum aestivum***

Saskia Maria van der Vies

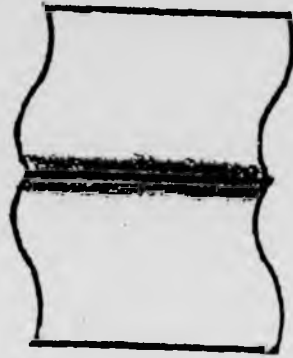
A thesis presented for the degree of Doctor of Philosophy

Department of Biological Sciences, University of Warwick

Coventry, United Kingdom

November 1989

**THIS VOLUME HAS A  
VERY TIGHT BINDING**



I would like to dedicate this thesis with much love to my parents

Dit proefschrift is met liefde opgedragen aan mijn ouders



The growth of science consists not in the addition of new certainties to existing certainties but in the replacement of existing theories by better theories - in the light of which any explanation that had been offered by an earlier theory might have to be revised.

Magee, B (1989) In: Science, Truth and Morals, The Weekend Guardian, 9th -10th of September.

## SUMMARY

This thesis describes the discovery of a new class of related proteins which has been named the chaperonins (Hemmingsen, Woolford, van der Vies, Tilly, Dennis, Georgopoulos, Hendrix & Ellis, Nature 333, 330-334, 1988). The proteins in this highly conserved class are structurally and immunologically related and ubiquitous in their occurrence in plastids, mitochondria and bacteria. The chaperonins comprise one class of the larger family of molecular chaperones since their function of assisting in the folding and assembly of other polypeptides without being components of the final structure meets the criteria suggested for molecular chaperones (Ellis, Nature 328, 378-379, 1987).

The chaperonin class of proteins was discovered during studies on the assembly of the hexadecameric enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco); this enzyme is found in the chloroplasts of plants where it catalyzes the first step in the pathways of both photosynthesis and photorespiration. The assembly of Rubisco *in vivo* had been proposed to require the activity of another chloroplast protein, originally known as the Rubisco large subunit binding protein because it binds to Rubisco large subunits newly-synthesized in isolated chloroplasts of *Pisum sativum* (Barraclough & Ellis, Biochim. Biophys. Acta 608, 19-31, 1980). Newly-imported Rubisco small subunits have since been shown to bind to the same chloroplast protein, which has therefore been renamed the Rubisco subunit binding protein (abbreviated to binding protein), or the plastid chaperonin.

Antibodies raised against the plastid chaperonin purified from *P. sativum* recognize two subunit polypeptides with an apparent Mr of 61 500 and 60 500 (termed alpha and beta respectively) in extracts of *Triticum aestivum* leaves. With the aid of these antibodies a cDNA fragment has been isolated and sequenced from a lambda gt11 expression library of cDNA from leaves of *Triticum aestivum*. The cDNA fragment of 1834 bp encodes the entire mature plastid chaperonin alpha subunit plus two amino acids of the presequence. The amino acid sequence of the *T. aestivum* alpha chaperonin shows 46% identity to a protein from *Escherichia coli* known as the *groEL* protein; this protein had previously been shown to be essential for cell viability and is required for the assembly of bacteriophage capsids. An identity of 59% is found between the wheat alpha chaperonin and a *groEL*-like protein present in *Mycobacterium leprae* and *M. tuberculosis*. Immunologically related proteins were also detected in a variety of prokaryotes including cyanobacteria and *Prochlorothrix hollandica*, the eukaryote *Chlamydomonas reinhardtii* and in mitochondrial fractions from *P. sativum* leaves and *Solanum tuberosum* tubers. All these related proteins comprise the new class of chaperonins.

Amino acid sequences of all the known chaperonins from plastids, mitochondria and bacteria were compared and show 41%-58% amino acid identity. The chloroplast alpha chaperonin is as closely related to the chloroplast beta chaperonin (amino acid identity is 50%), as it is to the bacterial and mitochondrial chaperonins. The most interesting finding to emerge from the analysis of the deduced amino acid sequences is the presence of a possible dinucleotide binding site. A highly conserved region of 36 amino acids shows 9 out of 11 matches reported for the dinucleotide binding site fingerprint (Wierenga, Terpstra & Holl, J. Mol. Biol. 187, 101-107, 1986), but only when the chaperonin sequence is read from the carboxy to the aminoterminal and two additional amino acids are allowed. Other proteins such as the  $\text{Ca}^{2+}$ -ATPase of the sarcoplasmic reticulum and the ecdysone-induced protein Eip 28/29 of *Drosophila melanogaster* also contain this reversed dinucleotide binding site sequence. This finding raises the novel possibility that a given binding site can be constructed from a set of amino acids running in either direction along the polypeptide chain; this possibility should be tested for other consensus sequences.

When the *T. aestivum* chloroplast alpha chaperonin is synthesized in *E. coli* cells, it forms a hybrid oligomeric complex with the host chaperonin. The *T. aestivum* Rubisco large subunits that are synthesized in *E. coli* are found associated with either the *E. coli* chaperonin or with the hybrid chaperonin complex, whereas co-synthesized Rubisco small subunits bind neither to the large subunits nor to the chaperonin complexes. The *T. aestivum* Rubisco subunits fail to assemble into an enzymically active oligomer when synthesized in the presence of the *T. aestivum* chloroplast alpha chaperonin. This work is discussed in light of the conclusion emerging from studies in several laboratories that chaperonins function in many processes within the cell, the common feature of which is the requirement to prevent folding and assembly occurring between transiently exposed interactive protein surfaces.

## SUMMARY

This thesis describes the discovery of a new class of related proteins which has been named the chaperonins (Hemmingsen, Woolford, van der Vies, Tilly, Dennis, Georgopoulos, Hendrix & Ellis, Nature 333, 330-334, 1988). The proteins in this highly conserved class are structurally and immunologically related and ubiquitous in their occurrence in plastids, mitochondria and bacteria. The chaperonins comprise one class of the larger family of molecular chaperones since their function of assisting in the folding and assembly of other polypeptides without being components of the final structure meets the criteria suggested for molecular chaperones (Ellis, Nature 328, 378-379, 1987).

The chaperonin class of proteins was discovered during studies on the assembly of the hexadecameric enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco); this enzyme is found in the chloroplasts of plants where it catalyzes the first step in the pathways of both photosynthesis and photorespiration. The assembly of Rubisco *in vivo* had been proposed to require the activity of another chloroplast protein, originally known as the Rubisco large subunit binding protein because it binds to Rubisco large subunits newly-synthesized in isolated chloroplasts of *Pisum sativum* (Barracough & Ellis, Biochim. Biophys. Acta 608, 19-31, 1980). Newly-imported Rubisco small subunits have since been shown to bind to the same chloroplast protein, which has therefore been renamed the Rubisco subunit binding protein (abbreviated to binding protein), or the plastid chaperonin.

Antibodies raised against the plastid chaperonin purified from *P. sativum* recognize two subunit polypeptides with an apparent Mr of 61 500 and 60 500 (termed alpha and beta respectively) in extracts of *Triticum aestivum* leaves. With the aid of these antibodies a cDNA fragment has been isolated and sequenced from a lambda gt11 expression library of cDNA from leaves of *Triticum aestivum*. The cDNA fragment of 1834 bp encodes the entire mature plastid chaperonin alpha subunit plus two amino acids of the presequence. The amino acid sequence of the *T. aestivum* alpha chaperonin shows 46% identity to a protein from *Escherichia coli* known as the *groEL* protein; this protein had previously been shown to be essential for cell viability and is required for the assembly of bacteriophage capsids. An identity of 59% is found between the wheat alpha chaperonin and a *groEL*-like protein present in *Mycobacterium leprae* and *M. tuberculosis*. Immunologically related proteins were also detected in a variety of prokaryotes including cyanobacteria and *Prochlorothrix hollandica*, the eukaryote *Chlamydomonas reinhardtii* and in mitochondrial fractions from *P. sativum* leaves and *Solarum tuberosum* tubers. All these related proteins comprise the new class of chaperonins.

Amino acid sequences of all the known chaperonins from plastids, mitochondria and bacteria were compared and show 41%-58% amino acid identity. The chloroplast alpha chaperonin is as closely related to the chloroplast beta chaperonin (amino acid identity is 50%), as it is to the bacterial and mitochondrial chaperonins. The most interesting finding to emerge from the analysis of the deduced amino acid sequences is the presence of a possible dinucleotide binding site. A highly conserved region of 36 amino acids shows 9 out of 11 matches reported for the dinucleotide binding site fingerprint (Wierenga, Terpstra & Holl, J. Mol. Biol. 187, 101-107, 1986), but only when the chaperonin sequence is read from the carboxy to the aminoterminal and two additional amino acids are allowed. Other proteins such as the  $\text{Ca}^{2+}$ -ATPase of the sarcoplasmic reticulum and the ecdysone-induced protein Eip 28/29 of *Drosophila melanogaster* also contain this reversed dinucleotide binding site sequence. This finding raises the novel possibility that a given binding site can be constructed from a set of amino acids running in either direction along the polypeptide chain; this possibility should be tested for other consensus sequences.

When the *T. aestivum* chloroplast alpha chaperonin is synthesized in *E. coli* cells, it forms a hybrid oligomeric complex with the host chaperonin. The *T. aestivum* Rubisco large subunits that are synthesized in *E. coli* are found associated with either the *E. coli* chaperonin or with the hybrid chaperonin complex, whereas co-synthesized Rubisco small subunits bind neither to the large subunits nor to the chaperonin complexes. The *T. aestivum* Rubisco subunits fail to assemble into an enzymically active oligomer when synthesized in the presence of the *T. aestivum* chloroplast alpha chaperonin. This work is discussed in light of the conclusion emerging from studies in several laboratories that chaperonins function in many processes within the cell, the common feature of which is the requirement to prevent folding and assembly occurring between transiently exposed interactive protein surfaces.

## ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor R. John Ellis F.R.S. for providing the opportunity to carry out this research project, and for his guidance, encouragement and valuable criticism during the course of this work. Thanks for the very pleasant and stimulating time, John!

Special thanks are due to Dr. Anthony A. Gatenby (E.I. du Pont de Nemours, Wilmington, USA) and to Dr. Sean M. Hemmingsen (Institute for Plant Biotechnology, Saskatoon, Canada) for their friendship and collaboration and the many fruitful discussions 'across the Atlantic'

I am indebted to Professor Richard B. Flavell for his support and for always allowing me to return to the Institute of Plant Science Research at Cambridge to perform 'yet another experiment'. Without the hospitality of Dr. Michael W. Bevan, and the support and help of Dr. Doug Bradley and Dr. George Murphy, nucleotide sequence determination and analysis would not have been so easy and enjoyable; with a special big thank you to George for guiding me through the mountain of computer programmes during my visits to the Institute for Plant Sciences Research at Cambridge

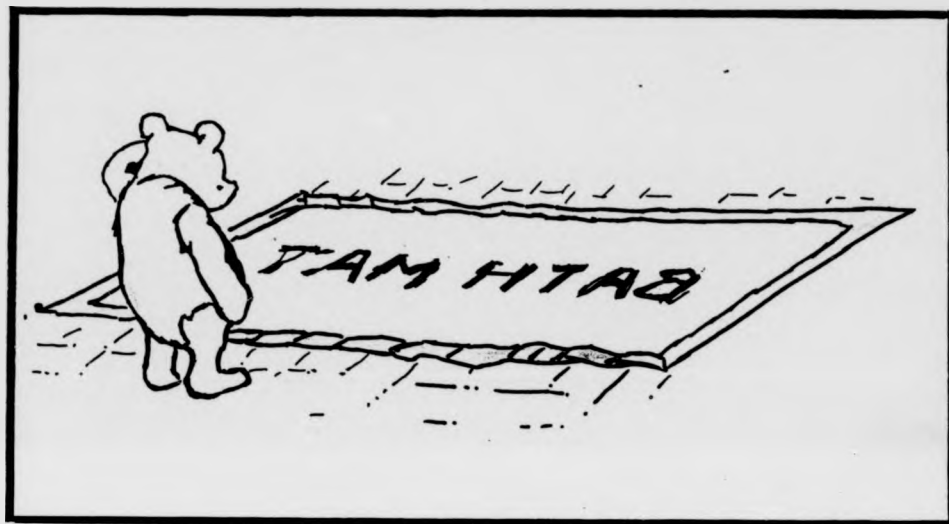
I am grateful to Dr. Chris J. Howe (Cambridge University) and his colleagues for their help and advice during my visit to the Department of Biochemistry where a series of immunological experiments was performed. The *Triticum aestivum* cDNA expression library in lambda gt 11 was a generous gift of Dr. Christine Raines and Dr. Tristan A. Dyer (Institute for Plant Science Research at Cambridge) who I would like to thank for providing instructions and advice on screening the library. I would like to express my thanks to Dr. Jan van Embden (Rijksinstituut voor Volksgezondheid en Milieuhygiene, Bilthoven, The Netherlands) for antibodies against the bacterial common antigen from *Pseudomonas aeruginosa*, to Professor Richard L. Hallberg (Iowa State University, Ames, USA) for providing antibodies against the mitochondrial chaperonin and helpful discussion, to Dr. Alfred J. Keys (Rothamsted Experimental Station, Harpenden) for his gift of purified *Triticum aestivum* chloroplast chaperonin, to Professor Chris J. Leaver F.R.S. and Andrew

Liddell (Edinburgh University) for providing mitochondrial fractions from *Pisum sativum* leaves and *Solanum tuberosum* tubers and for the pleasant collaboration, and to Dr. Angela Mehler (Hammersmith Hospital, London) for her gift of purified *Escherichia coli* chaperonin.

I would like to thank Dr. Johannes van der Vies (Pa!) for writing a computer programme that eased the chaperonin sequence analysis, and Dr. Andrew F.W. Coulson (Edinburgh University) for performing a number of database searches. The pleasant and fruitful discussions with Dr. George H. Lorimer (E.I. du Pont de Nemours, Wilmington, USA) and Dr. Douglas B. Young (Hammersmith Hospital, London) and the access to their unpublished data were much appreciated.

I would like to take the opportunity to thank the members of the Departments of Biological Sciences and Computer Services at the University of Warwick for their help and advice. I am grateful to Professor Roger Whittenbury and the University of Warwick for providing financial support to carry out this project, and to Marie-José van Neerven for the skillful typing of this thesis

Finally I would like to thank my parents, family and friends for their love and support.



Research is to see  
what everybody has seen  
and to think  
what nobody has thought

A. Szent-Györgyi

SUMMARY	4
DECLARATION	5
ACKNOWLEDGEMENTS	6
CONTENTS	9
ABBREVIATIONS	18
LIST OF FIGURES	21
LIST OF TABLES	26

## 1 LITERATURE REVIEW

1.1	INTRODUCTION	29
1.2	RIBULOSE-1.5-BISPHOSPHATE CARBOXYLASE/OXYGENASE	
1.2.1	<u>MOLECULAR STRUCTURE</u>	31
1.2.2	<u>MECHANISM OF CATALYSIS</u>	33
1.2.3	<u>SUBUNIT SYNTHESIS</u>	36
1.2.4	<u>SUBUNIT ASSEMBLY</u>	39
1.2.5	<u>GENE ORGANISATION AND EXPRESSION</u>	43
1.3	RUBISCO LARGE SUBUNIT BINDING PROTEIN	
1.3.1	<u>DISCOVERY</u>	47
1.3.2	<u>PURIFICATION</u>	48
1.3.3	<u>SYNTHESIS AND OCCURRENCE</u>	49
1.3.4	<u>DISSOCIATION BY M<sub>2</sub>ATP</u>	51
1.3.5	<u>THE RUBISCO LARGE SUBUNIT BINDING PROTEIN: - A MOLECULAR CHAPERONE?</u>	53
1.4	SYNTHESIS AND ASSEMBLY OF RUBISCO SUBUNITS IN <i>ESCHERICHIA COLI</i>	
1.4.1	<u>SYNTHESIS OF PLANT RUBISCO SUBUNITS</u>	55
1.4.2	<u>SYNTHESIS AND ASSEMBLY OF PROKARYOTIC</u>	

	<u>RUBISCO SUBUNITS</u>	58
1.4.3	<u>SYNTHESIS AND ASSEMBLY OF HETEROLOGOUS RUBISCO SUBUNITS</u>	60
1.5	AIM OF THE PRESENT WORK	61
2	MATERIALS AND METHODS	
2.1	MATERIALS	
2.1.1	<u>BIOLOGICAL MATERIALS</u>	63
2.1.1.1	Plants	63
2.1.1.2	Animals	63
2.1.2	<u>NON-BIOLOGICAL MATERIALS</u>	63
2.1.2.1	Chemicals and reagents	63
2.1.2.2	Enzymes	65
2.2	METHODS	
2.2.1	<u>GROWTH AND MAINTENANCE OF BIOLOGICAL MATERIALS</u>	67
2.2.1.1	Growth of plants	67
2.2.1.2	Growth and storage of bacterial strains	67
2.2.1.2.1	Growth of bacterial strains	67
2.2.1.2.2	Storage of bacterial strains	67
2.2.1.3	Growth and storage of bacteriophages	68
2.2.1.3.1	Growth of bacteriophages	68
2.2.1.3.2	Storage of bacteriophages	68
2.2.2	<u>MEDIA AND ANTIBIOTICS</u>	68
2.2.2.1	Media containing either agar or agarose	70
2.2.2.2	Antibiotic preparation	70
2.2.3	<u>BACTERIAL TRANSFORMATION</u>	70



2.2.4	<b><u>BACTERIAL TRANSFECTION</u></b>	71
2.2.4.1	<b>Bacteriophage <math>\lambda</math>.</b>	71
2.2.4.2	<b>Bacteriophage M13</b>	72
2.2.4.3	<b>High-titre lysate</b>	72
2.2.5	<b><u>NUCLEIC ACID ISOLATION</u></b>	73
2.2.5.1	<b>DNA extraction and precipitation</b>	73
2.2.5.1.1	DNA extraction with phenol/chloroform	73
2.2.5.1.2	DNA precipitation with ethanol	73
2.2.5.2	<b>Large-scale isolation of plasmid DNA</b>	74
2.2.5.2.1	Large-scale isolation of plasmid DNA on CsCl gradients	74
2.2.5.2.2	Rapid large-scale isolation of plasmid DNA	75
2.2.5.3	<b>Small scale plasmid DNA</b>	76
2.2.5.3.1	Rapid isolation of plasmid DNA	76
2.2.5.3.2	Isolation of highly purified plasmid DNA	77
2.2.5.4	<b>Bacteriophage <math>\lambda</math> DNA isolation</b>	78
2.2.6	<b><u>ISOLATION OF cDNA SEQUENCES</u></b>	79
2.2.6.1	<b>Screening of a <math>\lambda</math> gt 11 expression library with antibodies</b>	79
2.2.6.2	<b>Purification of plaques</b>	80
2.2.7	<b><u>NUCLEIC ACID RESTRICTION AND MODIFICATION</u></b>	81
2.2.7.1	<b>Restriction endonuclease digestion</b>	81
2.2.7.2	<b>Dephosphorylation of DNA</b>	81
2.2.7.3	<b>DNA ligation</b>	82
2.2.7.4	<b>Filling recessed 3' ends of double-stranded DNA</b>	83
2.2.7.5	<b>Removal of 5'mononucleotides from double-stranded DNA</b>	83
2.2.7.6	<b>Degradation of single-stranded DNA</b>	84
2.2.7.7	<b>Radiolabelling of DNA</b>	85
2.2.8	<b><u>ELECTRO-ELUTION OF DNA FROM AGAROSE GELS</u></b>	86

2.2.9	<u>BACTERIOPHAGE LAMBDA PLATE LYSATE PREPARATION</u>	86
2.2.10	<u>NUCLEIC ACID HYBRIDIZATION</u>	87
2.2.11	<u>DNA SEQUENCE DETERMINATION</u>	89
2.2.12	<u>GELELECTROPHORESIS</u>	91
2.2.12.1	<b>Nucleic acid analysis</b>	91
2.2.12.1.1	DNA analysis on high resolution agarose gels	91
2.2.12.1.2	Rapid DNA analysis on agarose gels	92
2.2.12.1.3	DNA analysis under denaturing conditions	92
2.2.12.1.4	RNA analysis on gels containing formaldehyde	94
2.2.12.2	<b>Protein analysis</b>	95
2.2.12.2.1	Protein analysis under non-denaturing conditions	95
2.2.12.2.2	Protein analysis under denaturing conditions	96
2.2.13	<u>PRODUCTION AND PURIFICATION OF ANTIBODIES</u>	98
2.2.13.1	<b>Production of antibodies</b>	98
2.2.13.2	<b>Purification of antibodies</b>	99
2.2.13.2.1	Purification of antibodies using a $\lambda$ gt 11 plate lysate	99
2.2.13.2.2	Purification of antibodies using a total <i>E.coli</i> protein extract	100
2.2.14	<u>IMMUNOBLOTTING OF PROTEINS AND ANTIBODY DETECTION</u>	101
2.2.14.1	<b>Immunoblotting of proteins</b>	101
2.2.14.2	<b>Antibody detection</b>	102
2.2.14.2.1	Enzyme-linked detection	102
2.2.14.2.2	Radiolabelled detection	103
2.2.15	<u>PREPARATION OF IODINATED PROTEIN A</u>	103
2.2.16	<u>PROTEIN DETECTION</u>	104
2.2.17	<u>TRANSCRIPTION AND TRANSLATION <i>IN VITRO</i></u>	105
2.2.17.1	<b>Transcription <i>in vitro</i></b>	105
2.2.17.2	<b>Translation <i>in vitro</i></b>	106
2.2.18	<u>IMPORT OF POLYPEPTIDES BY ISOLATED CHLOROPLASTS</u>	107

2.2.18.1	Isolation of chloroplasts	107
2.2.18.2	Import of polypeptides by chloroplasts	108
2.2.19	<u>SYNTHESIS OF FOREIGN PROTEINS IN <i>ESCHERICHIA COLI</i></u>	109
2.2.20	<u>PREPARATION OF A PROTEIN EXTRACT</u>	110
2.2.20.1	Preparation of a 'soluble' <i>Escherichia coli</i> protein extract	110
2.2.20.2	Preparation of a total plant leaf protein extract	110
2.2.21	<u>VELOCITY SEDIMENTATION CENTRIFUGATION</u>	111
2.2.22	<u>ENZYME ASSAYS</u>	111
2.2.22.1	Determination of Rubisco activity	111
2.2.22.2	Determination of $\beta$ -galactosidase activity	112
3	<b>RESULTS AND DISCUSSION</b>	
3.1	<b>ISOLATION AND CHARACTERIZATION OF cDNA SEQUENCES FROM <i>TRITICUM AESTIVUM</i></b>	114
3.1.1	<u>INTRODUCTION</u>	114
3.1.2	<u>CHARACTERIZATION AND ANALYSIS OF THE <i>PISUM SATIVUM</i> RUBISCO SUBUNIT BINDING PROTEIN ANTIBODIES</u>	114
3.1.3	<u>ISOLATION OF <i>TRITICUM AESTIVUM</i> cDNA SEQUENCES FROM THE LAMBDA GT 11 EXPRESSION LIBRARY</u>	116
3.1.4	<u>CHARACTERIZATION OF THE ISOLATED <i>TRITICUM AESTIVUM</i> cDNA SEQUENCE</u>	117
3.1.4.1	Subcloning of the cDNA sequence	117
3.1.4.2	Restriction enzyme digestion analysis of pSV8	118
3.1.4.3	DNA hybridization analysis of pSV8	118

3.1.4.4	Transcription and translation of the <i>Triticum aestivum</i> sequence <i>in vitro</i>	119
3.1.5	<u>DISCUSSION</u>	121
3.2	NUCLEOTIDE SEQUENCE DETERMINATION AND ANALYSIS OF THE ISOLATED <i>TRITICUM AESTIVUM</i> cDNA	
3.2.1	<u>INTRODUCTION</u>	124
3.2.2	<u>DETERMINATION OF THE <i>TRITICUM AESTIVUM</i> cDNA SEQUENCE</u>	124
3.2.2.1	Construction of pSV10 and pSV11	124
3.2.2.2	Generation of <i>Triticum aestivum</i> cDNA deletion derivatives	125
3.2.2.3	Nucleotide sequence determination	126
3.2.3	<u>ANALYSIS OF THE DNA SEQUENCE</u>	126
3.2.3.1	Identification of the translation initiation site	126
3.2.3.2	Analysis of the 3' non-coding region	127
3.2.3.3	Codon usage	129
3.2.4	<u>IDENTIFICATION OF THE TRANSLATED POLYPEPTIDE</u>	130
3.2.5	<u>ANALYSIS OF THE AMINO ACID SEQUENCE OF THE RUBISCO SUBUNIT BINDING PROTEIN ALPHA SUBUNIT</u>	131
3.2.5.1	Analysis of the amino acid composition	131
3.2.5.2	Determination of the hydropathy profile	131
3.2.5.3	Secondary structure prediction	132
3.2.5.4	Identification of the possible active sites and consensus binding sites	133
3.3.6	<u>DISCUSSION</u>	135
3.3.6.1	Identification of the <i>Triticum aestivum</i> cDNA fragment	135
3.3.6.2	The <i>Triticum aestivum</i> binding protein $\alpha$ subunit amino	

	acid sequence	137
3.3.6.3	The <i>Triticum aestivum</i> binding protein $\alpha$ subunit nucleotide sequence	139
3.2.6.4	Concluding remarks	142
3.3	IDENTIFICATION AND ANALYSIS OF PROTEINS RELATED TO THE RUBISCO SUBUNIT BINDING PROTEIN ALPHA SUBUNIT OF <i>TRITICUM AESTIVUM</i>	
3.3.1	<u>INTRODUCTION</u>	143
3.3.2	<u>PROTEINS RELATED IMMUNOLOGICALLY OR BY SEQUENCE TO THE RUBISCO SUBUNIT BINDING PROTEIN ALPHA SUBUNIT</u>	143
3.3.2.1	The <i>groEL</i> protein of <i>Escherichia coli</i>	143
3.3.2.2	The bacterial 65 kDa common antigen	145
3.3.2.3	The <i>groEL</i> -like protein in mitochondria of <i>Tetrahymena thermophila</i> and <i>Saccharomyces cerevisiae</i>	148
3.3.3	<u>IDENTIFICATION OF THE CHAPERONIN IN MITOCHONDRIA OF <i>SOLANUM TUBEROSUM</i> AND <i>PISUM SATIVUM</i></u>	150
3.3.4	<u>STRUCTURAL AND AMINO ACID SEQUENCE ANALYSIS OF THE CHAPERONIN CLASS</u>	153
3.3.4.1	Introduction	153
3.3.4.2	The plastid chaperonin group	154
3.3.4.3	The prokaryotic chaperonin group	155
3.3.4.4	The mitochondrial chaperonin group	156
3.3.4.5	The chaperonin class	157
3.3.5	<u>PROTEINS RELATED FUNCTIONALLY TO THE CHAPERONINS</u>	159

3.3.5.1	The <i>groES</i> protein from <i>Escherichia coli</i>	162
3.3.5.2	The 30S ribosomal A-protein from <i>Escherichia coli</i>	164
3.3.6	<u>IDENTIFICATION OF THE REVERSED DINUCLEOTIDE BINDING SITE SEQUENCE IN OTHER PROTEINS</u>	165
3.3.7	<u>DISCUSSION</u>	167
3.3.7.1	Chaperonins and oligomeric protein assembly	168
3.3.7.2	Chaperonins and protein transport	171
3.3.7.3	Chaperonins and nucleic acid metabolism	173
3.3.7.4	Chaperonins and protein synthesis	175
3.3.7.5	Chaperonins as stress proteins and immunogens	176
3.3.7.6	Chaperonin ultra structure and location	177
3.3.7.7	Chaperonins: mechanism of action	178
3.3.7.8	Concluding remarks	181
3.4	<b>EXPRESSION OF THE <i>TRITICUM AESTIVUM</i> ALPHA CHAPERONIN AND THE RUBISCO LARGE AND SMALL SUBUNITS IN <i>ESCHERICHIA COLI</i></b>	
3.4.1	<u>INTRODUCTION</u>	182
3.4.2	<u>EXPRESSION OF THE <i>TRITICUM AESTIVUM</i> CHLOROPLAST ALPHA CHAPERONIN IN <i>ESCHERICHIA COLI</i></u>	183
3.4.2.1	Construction of the expression plasmid pSV15	183
3.4.2.2	Synthesis of the <i>Triticum aestivum</i> chloroplast $\alpha$ chaperonin in <i>Escherichia coli</i>	184
3.4.2.3	Assembly of the <i>Triticum aestivum</i> chloroplast $\alpha$ chaperonin into homologous and heterologous oligomers in <i>Escherichia coli</i>	186
3.4.2.4	Analysis of the <i>Escherichia coli</i> and the <i>Triticum aestivum</i> chloroplast $\alpha$ chaperonins in <i>Escherichia coli</i>	

	extracts by velocity sedimentation centrifugation	188
3.4.3	<u>A COMPARISON OF THE FORM OF <i>TRITICUM AESTIVUM</i> RUBISCO SUBUNITS EXPRESSED IN <i>ESCHERICHIA COLI</i> WITH THE <i>PISUM SATIVUM</i> SUBUNITS IN ISOLATED CHLOROPLASTS</u>	190
3.4.4	<u>COEXPRESSION OF THE <i>TRITICUM AESTIVUM</i> CHLOROPLAST ALPHA CHAPERONIN AND THE RUBISCO LARGE AND SMALL SUBUNITS IN THE SAME <i>ESCHERICHIA COLI</i> CELL</u>	193
3.4.4.1	Construction of expression plasmid pSV16	193
3.4.4.2	Synthesis of the <i>Triticum aestivum</i> chloroplast $\alpha$ chaperonin and the Rubisco large and small subunits in the same <i>Escherichia coli</i> cell	194
3.4.5	<u>DISCUSSION</u>	195
4	CONCLUDING REMARKS	200
5	REFERENCES	204
6	INDEX	238

**HOMOLOGOUS PLANT AND BACTERIAL PROTEINS  
CHAPERONE OLIGOMERIC PROTEIN ASSEMBLY,  
Sean M. Hemmingsen, Carol Woolford, Saskia M. van  
der Vies, Kit Tilly, David T. Dennis, Costa P.  
Georgopoulos, Roger W. Hendrix & R. John Ellis,  
Nature 333, 330-334 (1988)**

## LIST OF ABBREVIATIONS

A	absorbance
APS	ammonium persulphate
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
CAB	chlorophyll <u>a/b</u> binding protein
cDNA	copy DNA
cf.	compare
CIP	calf intestine phosphatase
cpm	counts per minute
cps	counts per second
CsCl	cesium chloride
CTP	cytidine 5'-triphosphate
CY	casamino acid yeast medium
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
ddATP	2',3'-dideoxyadenosine 5'-triphosphate
ddCTP	2',3'-dideoxycytidine 5'-triphosphate
ddGTP	2',3'-dideoxyguanosine 5'-triphosphate
ddTTP	2',3'-dideoxythymidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTPs	2'-deoxynucleotide 5'-triphosphates
DTT	dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
EDTA	ethylenediaminetetra-acetic acid



EtBr	ethidium bromide
g	unit of gravitational field
GTP	guanosine 5'-triphosphate
h	hour
HCC	hexamine cobalt chloride
Hepes	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
IPTG	isopropyl- $\beta$ -D-thio-galactoside
kb	kilobases
kDa	kilo Dalton
LB	Luria-Bertani medium
mA	milliampere
MG	minimal glucose medium
min	minute
mRNA	messenger ribonucleic acid
Nr	relative molecular mass
NTP	nucleotide 5'-triphosphate
ONGP	O-nitro-phenol- $\beta$ -D-galactose pyranoside
p	plasmid
PBS	phosphate-buffered saline
PEG	polyethylene glycol
pfu	plaque forming unit
PMSF	phenylmethylsulphonyl fluoride
Poly(A)	polyadenylated
PVP	polyvinylpyrrolidone
RNA	ribonucleic acid
RNase	ribonuclease
RNasin	RNA inhibitor
Rubisco	ribulose-1,5-biphosphate carboxylase-oxygenase
RuBP	1.5-ribulose biphosphate

s	second
SDS	sodium dodecylsulphate
SIM	sucrose isolation medium
SM	sodium/magnesium phage buffer
SRM	sorbitol reaction medium
SSC	standard saline citrate buffer
ssDNA	single stranded DNA
SSPE	standard saline phosphate EDTA buffer
TAE	Tris-acetate-EDTA buffer
TBE	Tris-borate-EDTA buffer
TBS	Tris-buffered saline
TCA	trichloroacetic acid
TE	Tris-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
TET	Tris-EDTA-Triton buffer
T <sub>m</sub>	melting temperature
Tris	2-amino-2(hydroxymethyl)-1,3-propane diol
TS	Tris-sucrose buffer
U	unit
u.v.	ultraviolet
V	voltage
X-gal	5-bromo-4-chloro-3-indolyl- $\beta$ -galactoside
YT	yeast tryptone medium.

## LIST OF FIGURES

Figure 1. Photosynthetic and photorespiratory pathways.	29/1
Figure 2. Quaternary structure of Rubisco.	32/1
Figure 3. Rubisco catalysis and substrate binding.	35/1
Figure 4. Synthesis and assembly of hexadecameric Rubisco.	39/1
Figure 5. Immunological detection of the Rubisco subunit binding protein in extracts of <i>Triticum aestivum</i> .	115/1
Figure 6. Analysis of the purified Rubisco subunit binding protein antibody preparation.	115/3
Figure 7. Detection of the Rubisco subunit binding protein in $\lambda$ gt11 lysogens using antibodies.	117/1
Figure 8. Analysis of the <i>Triticum aestivum</i> cDNA from the lysogen SM8.	117/3
Figure 9. Determination of the orientation of the <i>Triticum aestivum</i> cDNA in the pGEM3 vector.	118/1
Figure 10. Structure of the plasmids pSV8 and pSV9.	118/3
Figure 11. Determination of the restriction enzyme sites in the <i>Triticum</i> <i>aestivum</i> cDNA sequence.	118/5
Figure 12. DNA hybridisation of the <i>Triticum aestivum</i> and the <i>Ricinus</i> <i>communis</i> cDNA sequences.	119/1
Figure 13. Transcription <i>in vitro</i> of the <i>Triticum aestivum</i> cDNA sequence.	120/1
Figure 14. Analysis of translation products from pSV8 and pSV9 transcripts.	120/3
Figure 15. Structure of plasmids pSV10 and pSV11.	125/1
Figure 16. Outline of the strategy for targeting deletions.	126/1
Figure 17. Timecourse analysis of pSV11 deletions.	126/3
Figure 18. Analysis of pSV11 deletion transformants.	126/5
Figure 19. DNA sequence determination of the 3' end containing the poly(A) tail.	126/7

Figure 20. Nucleotide and deduced amino acid sequence of the <i>Triticum aestivum</i> cDNA.	126/9
Figure 21. The 3' non-coding region of the cDNA from <i>Triticum aestivum</i> .	127/1
Figure 22. Identification of direct repeats in the <i>Triticum aestivum</i> cDNA 3' non-coding region.	128/1
Figure 23. Identification of self-complementary sequences in the <i>Triticum aestivum</i> cDNA 3' non-coding region.	129/1
Figure 24. Comparison of the aminoterminal sequences of the Rubisco subunit binding protein subunits from <i>Triticum aestivum</i> and <i>Pisum sativum</i> .	131/1
Figure 25. Hydropathy profile of the Rubisco subunit binding protein $\alpha$ subunit.	132/1
Figure 26. Secondary structure prediction of the Rubisco subunit binding protein $\alpha$ subunit.	132/3
Figure 27. A potential reversed dinucleotide binding sequence.	134/1
Figure 28. Antigenic profile of the Rubisco subunit binding protein $\alpha$ subunit.	135/1
Figure 29. Amino acid sequence comparison of the <i>Triticum aestivum</i> Rubisco binding protein $\alpha$ subunit and the <i>Escherichia coli</i> groEL protein.	144/1
Figure 30. Assembly pathway of bacteriophage lambda prohead.	145/1
Figure 31. Amino acid sequence comparison of the <i>Triticum aestivum</i> Rubisco subunit binding protein $\alpha$ subunit and the <i>Mycobacterium tuberculosis</i> 65 kDa antigen.	145/3
Figure 32. Immunologically related polypeptides to the <i>Pisum sativum</i> Rubisco subunit binding protein in prokaryotes.	146/1
Figure 33. Immunologically related polypeptides to the <i>Pseudomonas aeruginosa</i> common antigen.	146/3
Figure 34. Immunologically related polypeptides to the <i>Escherichia coli</i>	

<i>groEL</i> protein.	147/1
Figure 35. Cross reactivity of the antibodies to the <i>Pisum sativum</i> Rubisco subunit binding protein with polypeptides in prokaryotic cell extracts.	147/3
Figure 36. Cross reactivity of the antibodies to the purified $\alpha$ and $\beta$ subunits of the Rubisco subunit binding protein from <i>Pisum sativum</i> with prokaryotic proteins.	147/5
Figure 37. Immunologically related proteins in the cyanobacterium <i>Phormidium laminosum</i> .	148/1
Figure 38. Amino acid sequence comparison of the <i>Triticum aestivum</i> Rubisco subunit binding protein $\alpha$ subunit and the <i>Anacystis nidulans groEL</i> -like protein.	148/3
Figure 39. Immunological relationship between a mitochondrial <i>groEL</i> -like protein from <i>Tetrahymena thermophila</i> and the Rubisco subunit binding protein from <i>Pisum sativum</i> .	149/1
Figure 40. Immunologically related proteins to the <i>groEL</i> -like protein from <i>Tetrahymena thermophila</i> .	149/3
Figure 41. Identification of the chaperonin in isolated mitochondria from <i>Solarum tuberosum</i> .	151/1
Figure 42. Identification of the chaperonin in isolated mitochondria from <i>Pisum sativum</i> .	151/4
Figure 43. Release of extrinsic proteins from mitochondrial membranes isolated from <i>Solarum tuberosum</i> .	152/1
Figure 44. Release of extrinsic proteins from mitochondrial membranes isolated from <i>Pisum sativum</i> .	152/3
Figure 45. Amino acid sequence comparison of the plastid chaperonins.	154/3
Figure 46. Secondary structure predictions of the <i>Brassica napus</i> plastid chaperonin.	155/1
Figure 47. Amino acid sequence comparison of the prokaryotic chaperonins.	155/3

Figure 48. Comparison of secondary structure predictions of plastid and prokaryotic chaperonins.	156/1
Figure 49. Secondary structure predictions of the prokaryotic chaperonins.	156/4
Figure 50. Amino acid sequence comparison of the mitochondrial chaperonins.	156/7
Figure 51. Secondary structure predictions of the mitochondrial chaperonins.	156/10
Figure 52. Amino acid sequence comparison of the chaperonins.	157/1
Figure 53. The reversed dinucleotide binding site sequence and its secondary structure prediction.	157/6
Figure 54. Amino acid sequence comparison of the <i>groES</i> -like proteins.	162/3
Figure 55. Representation of the amino acid comparisons of <i>groES</i> -like proteins and the chaperonins.	162/5
Figure 56. Secondary structure predictions of the <i>groES</i> -like proteins.	163/3
Figure 57. Amino acid sequence comparison of the 30S ribosomal A-protein from <i>E.coli</i> and a number of chaperonins.	164/1
Figure 58. A reversed dinucleotide binding site in the (Ca <sup>2+</sup> -Mg <sup>2+</sup> )-dependent ATPase and the ecdysome inducible protein.	165/1
Figure 59. Structure of the chaperonin from <i>Pisum sativum</i> .	177/1
Figure 60. Nucleotide sequence determination of the translational fusion region in pSV15.	184/1
Figure 61. Structure of expression plasmid pSV15.	184/3
Figure 62. Synthesis of the <i>Triticum aestivum</i> chloroplast $\alpha$ chaperonin in <i>E.coli</i> .	186/1
Figure 63. Analysis of <i>Triticum aestivum</i> chloroplast $\alpha$ chaperonin in <i>E.coli</i> extracts under non-denaturing conditions.	186/3
Figure 64. Analysis of <i>Triticum aestivum</i> chloroplast $\alpha$ chaperonin polypeptides in <i>E.coli</i> extracts by velocity sedimentation centrifugation in the absence of Mg <sup>2+</sup> ions.	188/1
Figure 65. Analysis of <i>Triticum aestivum</i> chloroplast $\alpha$ chaperonin	

polypeptides in <i>E.coli</i> extracts by velocity sedimentation centrifugation in the presence of $Mg^{2+}$ ions.	188/3
Figure 66. Association of imported Rubisco small subunits with the chloroplast chaperonin.	190/1
Figure 67. Structure of the <i>Triticum aestivum</i> Rubisco subunit expression vectors.	191/1
Figure 68. Association of <i>Triticum aestivum</i> Rubisco large subunits with the <i>E.coli</i> chaperonin in <i>E.coli</i> extracts.	191/3
Figure 69. Analysis of <i>Triticum aestivum</i> Rubisco small subunits in <i>E.coli</i> extracts by sucrose density gradient centrifugation.	192/1
Figure 70. Construction of the expression plasmid pSV16.	194/1
Figure 71. Analysis of the <i>Triticum aestivum</i> chloroplast $\alpha$ chaperonin and the Rubisco large and small subunits in the same <i>E.coli</i> cell extract.	194/3
Figure 72. A proposed model for chaperonin activity.	203/1

## LIST OF TABLES

Table 1.	DNA hybridization parameters	88/1
Table 2.	Composition of dideoxynucleotide mixes	90/1
Table 3.	Composition of non-denaturing protein gel mixes	95/1
Table 4.	Composition of SDS denaturing protein gel mixes	97/1
Table 5.	Sequences surrounding the AUG initiation codon	127/3
Table 6.	Preferences and avoidances of nucleotides at positions flanking the natural termination codons	128/3
Table 7.	Nucleotide composition of pooled protein encoded sequences	129/3
Table 8.	<i>Triticum aestivum</i> cDNA nucleotide composition	129/4
Table 9.	Coding usage of the <i>Triticum aestivum</i> cDNA coding region and of pooled protein encoding sequences from a variety of species	129/5
Table 10.	Percentages (G+C) content in protein encoded sequences from different species	129/10
Table 11.	Amino acid compositions of the mature <i>Triticum aestivum</i> $\alpha$ subunit and pooled protein sequences from different species	131/3
Table 12.	Active sites and consensus binding site sequences	133/1
Table 13.	Distribution of termination codons	139/1
Table 14.	Nucleotide frequencies at positions surrounding the AAUAAA- like motif	140/1
Table 15.	Percentage identical amino acids of the chaperonin proteins	154/1
Table 16.	Percentage conserved amino acids of the chaperonin proteins	155/7
Table 17.	Proteins possibly related to the chaperonins selected from the EMBL nucleotide database	160/1
Table 18.	Possibly chaperonin related proteins selected from the NBRF protein database	160/3
Table 19.	Percentage identical and conserved amino acids between the <i>groES</i> -like proteins and the chaperonin sequences	162/1



Table 20. Percentage identical and conserved amino acids of the *groES*-like protein

163/1

Table 21. Molecular chaperones

201/1

## 1 LITERATURE REVIEW

## 1.1 INTRODUCTION

The process whereby light energy is converted into chemical energy which is subsequently used to synthesize organic compounds is known as photosynthesis. Photosynthesis is the source of carbon and energy for most forms of life on earth and is undertaken by a number of groups of organisms including all green plants, the eukaryotic algae, diatoms, cyanobacteria and the purple bacteria. Photosynthesis can be divided into two phases:-

1. A light-dependent phase in which energy is used to generate a proton gradient which is used to synthesize ATP, and to produce reducing power in the form of NADPH at the expense of water oxidation.

2. A light-independent phase in which the reducing power and the ATP generated in the first phase are used for the fixation of carbon dioxide and its conversion to organic compounds.

In eukaryotic organisms the photosynthetic reactions take place in the chloroplast, and the fixation of carbon dioxide ( $\text{CO}_2$ ) occurs in the stromal compartment of the organelle in a series of enzyme-catalyzed reactions known as the Calvin cycle. The first step in this cycle is catalyzed by the enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco EC 4.1.1.39; Fig. 1). This bifunctional enzyme catalyzes either the carboxylation of the compound ribulose-1,5-bisphosphate (RuBP) using  $\text{CO}_2$  and  $\text{H}_2\text{O}$  to produce two molecules of 3-phosphoglycerate (PGA), or alternatively through the oxygenase reaction using one molecule of oxygen ( $\text{O}_2$ ) rather than  $\text{CO}_2$ , to produce one molecule of 3-phosphoglycerate and one molecule of 2-phosphoglycolate. The molecule 2-phosphoglycolate is the substrate for a series of reactions that lead to the loss of carbon as carbon dioxide and this process is referred to as photorespiration (Fig. 1; for review see Ogren, 1984). The efficiency with which carbon is fixed relative to oxygen determines the growth of the organism and is determined by the properties of the Rubisco enzyme and by the relative concentrations of  $\text{CO}_2$  and  $\text{O}_2$  around the enzyme.

The two pathways are presented schematically. Note that because the two complementary routes of glycine metabolism in the mitochondria, two molecules of 2-phosphoglycolate must enter the respiratory pathway for each molecule of serine, CO<sub>2</sub> and NH<sub>3</sub> by the glycine decarboxylase reaction. Symbols: RuBP, ribulose-1.5-bisphosphate; PGA, 3-phosphoglycerate; P-glycolate, 2-phosphoglycolate; OG, α-keto-oxoglutarate; OH-pyruvate, hydroxypyruvate; THF, tetrahydrofolic acid. Amino acids are shown in the three-letter code as described by IUPAC-IUB (1970;1984). This Figure was taken from a paper by Somerville *et al.* (1983).



Laing *et al.* (1974) showed that oxygen is linearly competitive for the carboxylase reaction and that carbon dioxide is linearly competitive for the oxygenase reaction. Since the early observation by Saussure in 1804 (cited in Kimball, 1983) that *Pisum sativum* plants exposed to high CO<sub>2</sub> concentrations grew better than control plants in ambient air, the effect of increased CO<sub>2</sub> concentration on agricultural yield has been studied intensively. Recent studies have indicated that with a doubling of the atmospheric CO<sub>2</sub> concentration the economic yield of many plants will probably increase by an average 33% (Kimball, 1983). Thus if it were possible to improve the efficiency with which CO<sub>2</sub> competes with O<sub>2</sub> at the active site of the Rubisco enzyme, the rate of photosynthesis would increase relative to the rate of photorespiration, with subsequent increases in crop yield (Ellis & Gatenby, 1984).

Evolution has produced a means of reducing photorespiration in some species, but always by means of mechanisms that increase the concentration of carbon dioxide around the active site of the enzyme, such as by pumping CO<sub>2</sub> across membranes which occurs in unicellular algae and cyanobacteria (Badger *et al.*, 1980; Spalding *et al.*, 1983), and by the C<sub>4</sub> type of photosynthesis found in some families of higher plants (Edwards & Walker, 1983). However the observed natural variation of the kinetic properties of the Rubisco enzyme from different sources indicates that modifications improving the efficiency of carboxylation have also occurred during evolution (Jordan & Ogren, 1981; 1983). It is because of this latter observation that Rubisco is a major target for genetic engineering to investigate whether it is possible to accelerate these evolutionary changes for the benefit of agriculture (Miziorko & Lorimer, 1983; Ellis & Gatenby, 1984).

Attempts to produce altered form of Rubisco by mutagenizing plant cells in order to prevent photorespiration have been unsuccessful so far (reviewed by Somerville & Ogren, 1982). In view of the difficulties in obtaining useful mutants, attempts have been made to express plant DNA in bacteria with the aim of assembling a functional Rubisco enzyme, that could provide the basis for mutagenic studies to alter the properties of the enzyme. These attempts have so far not succeeded in producing an active enzyme from DNA sequences coding for the Rubisco subunits from higher plants (Bradley *et al.*, 1986; Gatenby *et al.*, 1987), but have been successful in producing an active Rubisco enzyme using

cyanobacterial sequences. The synthesis and assembly of Rubisco subunits in *Escherichia coli* (*E.coli*) is reviewed in detail in Section 1.4.

The synthesis and assembly of the hexadecameric Rubisco enzyme in higher plants involves light and the coordinated interaction of two distinct genetic systems present in the chloroplast and the nucleus. The understanding of this clearly complex process has been of continuing interest to scientists and the next sections review and discuss some of the research which has contributed to our understanding of the synthesis and assembly of the Rubisco enzyme (Section 1.2.3 and Section 1.2.4), the regulation of gene expression (Section 1.2.5) and the structural (Section 1.2.1) and catalytic properties (Section 1.2.2) of the enzyme. The failure to assemble an active higher plant Rubisco enzyme from its subunits synthesized in *E.coli*, coupled with the observations made by Barraclough & Ellis (1980) have led to the proposal that another protein may be required for the assembly of Rubisco in plants. The discovery and the properties of this protein, called the Rubisco large subunit binding protein, are reviewed in Section 1.3.

The aim of this review is to describe and discuss observations that have led to the hypothesis that the co-synthesis of the Rubisco large subunit binding protein in the same *E.coli* cell that synthesizes the Rubisco subunits may be a way of rescuing the assembly of a functional enzyme, and thus allowing mutagenic studies to proceed.

## **1.2 RIBULOSE-1.5-BISPHOSPHATE CARBOXYLASE/OXYGENASE**

### **1.2.1 MOLECULAR STRUCTURE**

The Rubisco enzyme occurs in two distinct forms, which are referred to as form I and form II. The form I Rubisco is the most common and is found in some photosynthetic prokaryotes and all eukaryotes examined. The enzyme has a complex hexadecameric structure which contains eight large subunits with a Mr of 50 000 to 55 000 and eight small subunits with a Mr of 12 000 to 18 000, to give a total Mr of approximately 550 000. In contrast the form II Rubisco is an oligomer of large subunits only and found in non-sulphur

photosynthetic bacteria. In *Rhodospirillum rubrum* the large subunits are arranged in a dimeric form (Tabita & McFadden, 1974), whereas in *Rhodopseudomonas sphaeroides* a larger oligomer of probably six subunits has been identified (Gibson & Tabita, 1977)

Initial studies suggested that in the form I Rubisco enzyme from *Nicotiana tabacum* the large subunits are arranged as two layers of four subunits (reviewed by Miziorko & Lorimer, 1983). Recently however it has been shown by Chapman *et al.* (1986; 1987), who analyzed crystals of non-activated Rubisco from *N. tabacum* by X-ray diffraction at near-atomic resolution (3Å), that all large subunits are elongated along a four-fold axis, thus the molecule cannot simply be described as layers of subunits. The arrangement of the subunits in the Rubisco enzyme from *N. tabacum* as described by Chapman *et al.* (1986) is schematically shown in Figure 2A.

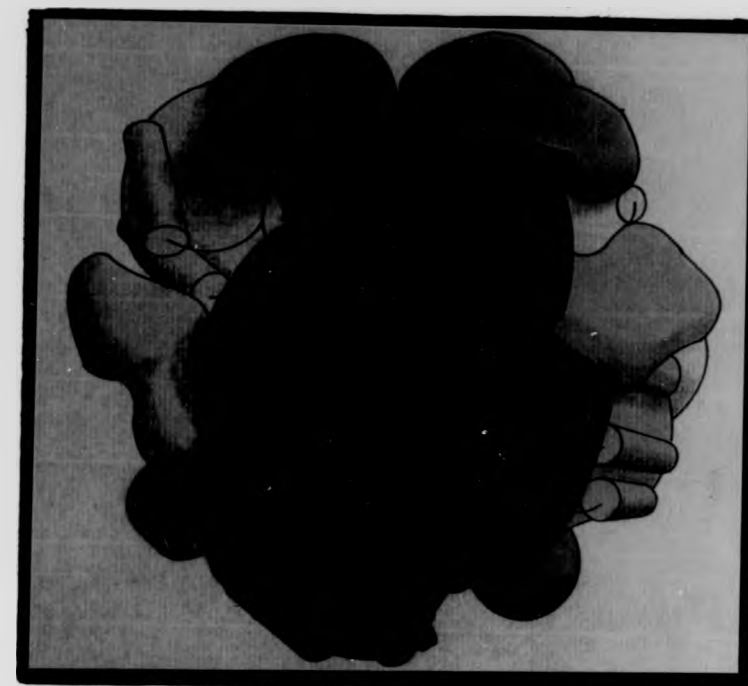
The small subunits are arranged in tetramers that are clustered around the four-fold axis at both the top and the bottom of the large subunit core structure, so that each large subunit lies between a pair of small subunits. Along the four-fold axis is a channel of 6-28Å, which had also been observed earlier in electron microscopic studies and which is accessible to solvent molecules (Eisenberg *et al.*, 1987). The level of resolution allows the identification of domains and main secondary structures. Each subunit contains two domains, one smaller N-terminal domain of approximately 150 residues and a larger C-terminal domain that has an  $\alpha/\beta$  barrel structure (Fig. 2A). Similar  $\alpha/\beta$  barrel motifs have also been observed in a number of other functionally different and genetically unrelated enzymes (reviewed by Schneider *et al.*, 1986) and seem to provide a basic framework for a number of biological functions. The active site of the *N. tabacum* Rubisco enzyme is probably at the interface of the large subunits at the opening of the  $\alpha/\beta$  barrel, facing the solution (Chapman *et al.*, 1987).

The large subunit of the form I enzyme from a range of different species are closely related and show a high percentage of amino acid sequence similarity ranging from 70% to 90% (Miziorko & Lorimer, 1983). In contrast the amino acid sequence similarity between the form I and the form II large subunits is only about 25% (Hartman *et al.*, 1982; Nargang *et al.*, 1984), but certain regions of high amino acid identity are conserved in all Rubisco

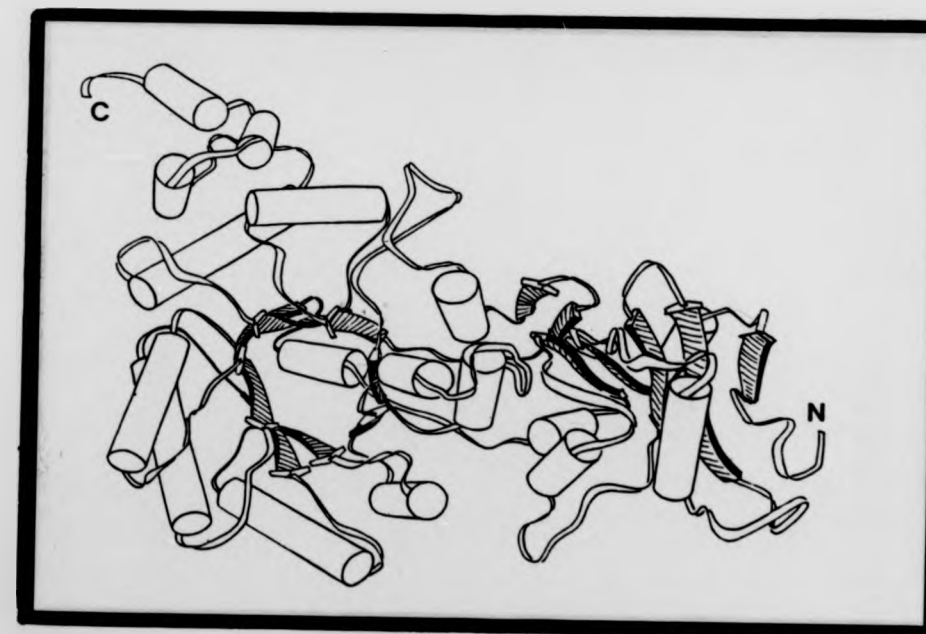
Figure 2. Quaternary structure of Rubisco

The spacial arrangement of the Rubisco subunits and domains in the Rubisco enzyme from *Nicotiana tabacum* is shown (Panel A). The small subunits are in blue, and each large subunit is in a different colour. Notice that the aminoterminal domain of the green large subunit is positioned close to the opening of the  $\alpha/\beta$  barrel of the red domain. This picture was taken from a paper by Eisenberg *et al.* (1987). A schematic diagram of the structure of one subunit of the Rubisco enzyme from *Rhodospirillum rubrum* is shown in Panel B, in which cylinders represent  $\alpha$ -helicies and arrows represent  $\beta$ -strands. This picture was taken from a paper by Schneider *et al.* (1986).

A



B





enzymes (some are reviewed by Gutteridge & Gatenby, 1987). The small subunits in the different form I Rubisco enzymes show a lower percentage of amino acid similarity than their large subunit counterparts, but also show regions of high amino acid identity which may be involved in the association with the large subunits.

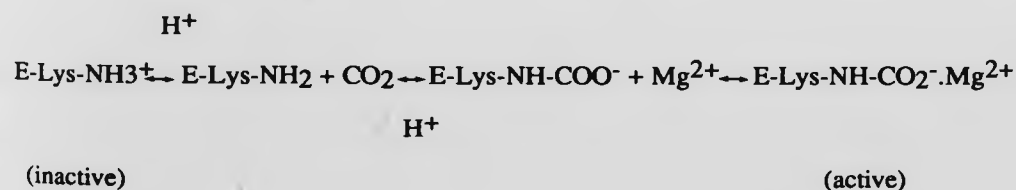
The form II enzyme from *Rhodospirillum rubrum* is a large subunit dimer and represents the most simple form of Rubisco that has been identified. The quaternary structure of the *R. rubrum* dimer has been determined at 2.9Å resolution by X-ray crystallography (Schneider *et al.*, 1986). The dimeric enzyme has the shape of an elongated cylinder with dimensions of 50Å x 75Å x 105Å. Each large subunit is made up of two domains and resembles the large subunit in the form I Rubisco from *N. tabacum* (Fig. 2B). The N-terminal and C-terminal domain interactions are few, however the interactions between the N-terminal domain of one subunit and the barrel structure of the second subunit are extensive. As in the form I Rubisco from *N. tabacum*, the active site in the *R. rubrum* dimer is formed by residues present on both the subunits. The quaternary structure of the *N. tabacum* Rubisco can thus be seen as the tetramerization of the *R. rubrum* large subunit dimers with the addition of small subunits, whose association with the large subunits may promote stabilization of the four dimers. The quaternary structures of the different Rubisco forms have implications for the assembly and the function of the plant enzyme (see also Section 1.3 and Section 1.4).

### 1.2.2 MECHANISM OF CATALYSIS

In a soluble extract of photosynthetic eukaryotic cells, up to 65% of the protein can be accounted for by the Rubisco enzyme, and thus Rubisco has been claimed to be 'the most abundant protein on earth' (Ellis, 1979). The reason for its abundance appears to be the fact that the enzyme has a low turnover number (t.o.). At saturating substrate concentrations one molecule of Rubisco fixes approximately 1000 molecules of CO<sub>2</sub> per minute (compare e.g. β-galactosidase, t.o.=12 500 and carbonic anhydrase, t.o.=3 6000 000 (Lehninger, 1970)). In the atmospheric concentrations of CO<sub>2</sub> the turnover number of the Rubisco enzyme is

about 200; thus the organism has to synthesize a large number of molecules to meet the required demand for carbon fixation. A low specificity (about  $2\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ ) has been determined for Rubisco purified from both prokaryotic and eukaryotic cells. It appears that even prokaryotes, which are under severe selection pressure, have not managed to evolve a more efficient Rubisco enzyme (Ellis & Gatenby, 1984). The abundance of Rubisco and its low specific activity make it an ideal candidate to study the molecular mechanism of catalysis which may elucidate why Rubisco is such an inefficient catalyst. The remainder of this section discusses the processes involved in catalysis in detail.

Pon *et al.* (1963) were the first to recognize that Rubisco must be activated before it catalyzes either the carboxylation or the oxygenation of ribulose-1,5-bisphosphate. Early studies showed that in both isolated intact chloroplasts and leaves the  $K_m(\text{CO}_2)$  was similar to the  $\text{CO}_2$  concentration ( $10\mu\text{M}$ ) found in aqueous solution in equilibrium with air, whereas the  $K_m(\text{CO}_2)$  for the Rubisco purified from *Spinacia oleracea* was approximately ten times higher (reviewed by Walker, 1973). Bahr & Jensen (1974) identified a form of the enzyme which has a high affinity for  $\text{CO}_2$ , in hypotonic preparations of lysed chloroplasts. This form of Rubisco was shown to be stabilized by  $\text{CO}_2$  (or bicarbonate) and  $\text{Mg}^{2+}$  ions (Bahr & Jensen, 1974; Andrews *et al.*, 1975). Thus  $\text{CO}_2$  acts both as an activator and a substrate. The activation of the Rubisco enzyme is an ordered process (Lorimer *et al.*, 1976) in which  $\text{CO}_2$  reacts with an  $\epsilon$ -lysyl amino group to form a carbamate, thereby generating a negatively-charged carbonyl group which provides a site for interaction with the positively-charged  $\text{Mg}^{2+}$  ion. Thus the activation of the enzyme can be described as follows (Lorimer & Miziorko, 1980), where E represents the Rubisco enzyme:-



The rate-limiting step appears to be the formation of the carbamate which occurs at Lys residue 201 in the spinach enzyme (Lorimer, 1981) and at Lys residue 91 in the *R. rubrum* dimeric enzyme (Donnelly *et al.*, 1983). Since the activation of Rubisco *in vivo* is promoted by light this raises the question how the observations made *in vitro* can be interpreted for the *in vivo* situation. It appears that during illumination the concentration of CO<sub>2</sub> probably changes little, whereas the Mg<sup>2+</sup> ion concentration and the pH in the stroma of the chloroplast increase (reviewed by Buchanan, 1980), and this could thus promote the carbamylation of the ε-amino group. However the K<sub>act</sub>(CO<sub>2</sub>) for Rubisco determined *in vitro* is approximately 25-30 μM (Lorimer *et al.*, 1976; Jordan & Ogren, 1981), whereas almost fully activated Rubisco has been observed *in vivo* at high light intensity in air, conditions under which the CO<sub>2</sub> concentration in the chloroplast is approximately 10 μM (Perchorowicz & Jensen, 1983; Salvucci *et al.*, 1986). These observations suggest that an additional factor which allows maximum activation of the Rubisco enzyme is present *in vivo*. With the aid of a mutant of *Arabidopsis thaliana* which was defective in Rubisco activation (Somerville *et al.*, 1982), the additional factor was identified as a soluble protein and was called Rubisco activase (Salvucci *et al.*, 1985). The Rubisco activase has been purified from *Spinacia oleracea* and activates purified Rubisco in an ATP-dependent fashion (Streusand & Portis, 1987) indicating that ATP hydrolysis may be required.

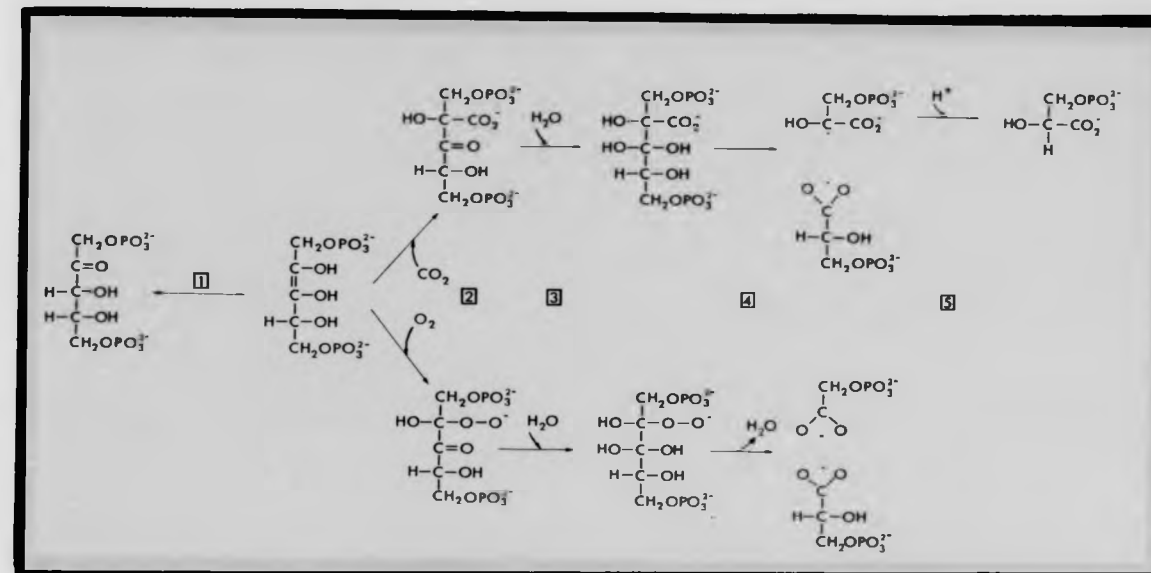
The reactions catalyzed by the activated Rubisco enzyme have been studied in great detail (reviewed by Lorimer *et al.*, 1986; Gutteridge & Gatenby, 1987). The fate of the various atoms involved in the formation of 3-phosphoglycerate and 2-phosphoglycolate from ribulose-1,5-bisphosphate and CO<sub>2</sub> and O<sub>2</sub> respectively is summarized in Figure 3. The first step is the formation of the enediol form of ribulose bisphosphate (product of reaction 1, Fig. 3A) which requires only the activated Rubisco enzyme. It is the enediol form to which either CO<sub>2</sub> or O<sub>2</sub> binds. Once the substrate is bound the RuBP is converted into either two molecules of 3-phosphoglycerate or one molecule of 3-phosphoglycerate and one molecule of 2-phosphoglycolate respectively. The recent determination of the quaternary structure of the dimeric Rubisco from *R. rubrum* at 1.7 Å resolution has provided detailed information about the mode of binding of CO<sub>2</sub>, the Mg<sup>2+</sup> ion and the transition state

Figure 3. Rubisco catalysis and substrate binding

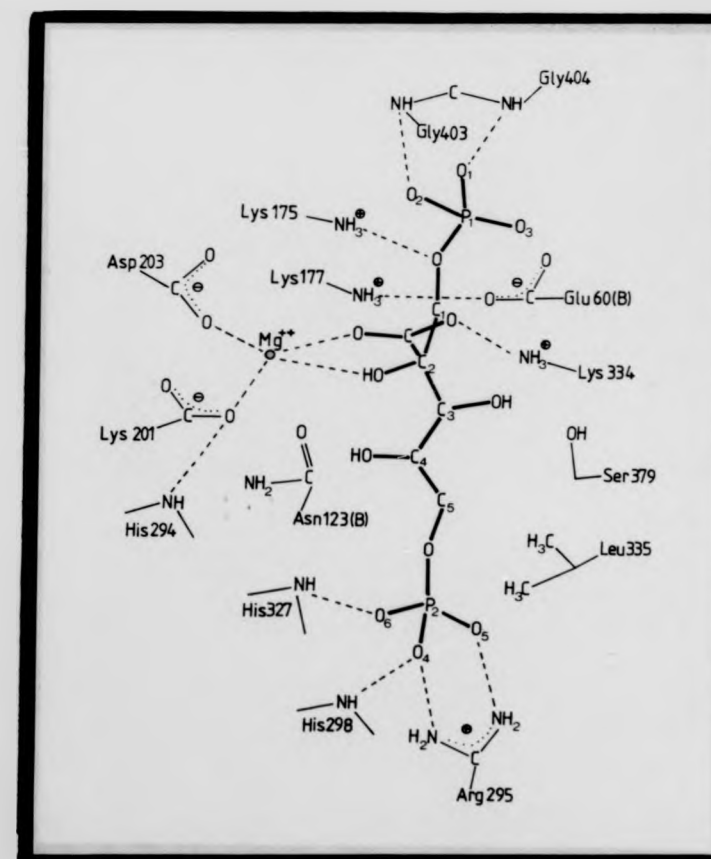
The overall reaction which is catalyzed by the Rubisco enzyme can be dissected into five steps (Panel A): 1. enolization, involving the protonation of C4 of ribulose-1,5-bisphosphate; 2. either initiation of carboxylation or oxygenation by electrophilic addition of gaseous carbon dioxide ( $\text{CO}_2$ ) or oxygen ( $\text{O}_2$ ) respectively; 3. hydration of the C3 carbonyl by  $\text{H}_2\text{O}$ ; 4. carbon-carbon cleavage to yield a molecule of 3-phosphoglycerate (lower molecule) from the bottom of the six carbon bisphosphate intermediate, and either one molecule of C-2 carbanion derivative or one molecule of 2-phosphoglycolate in the carboxylase and oxygenase reaction respectively; 5. protonation of the C-2 carbanion derivative on the enzyme to give the second molecule of 3-phosphoglycerate in the carboxylase reaction. This diagram was designed based on data by Lorimer *et al.* (1986) and Gutteridge & Gatenby (1987).

Panel B shows a schematic diagram of interactions between the Rubisco protein side chains and the metal atom (red) of the activated enzyme from *Spinacia oleracea*. The bound transition state analogue 2-carboxy-D-arabinitol-1,5-bisphosphate (2CABP) is shown in yellow. This diagram was taken from a recent paper by Andersson *et al.* (1989).

A



B



analogue 2-carboxy-D-arabinitol-1.5-bisphosphate (2CABP) to amino acid residues at the active site of the enzyme from *R.rubrum* and *S.oleracea* (Andersson *et al.*, 1989). The schematic diagram representing the different interactions is shown in Figure 3B.

Despite all the information available regarding the reactions catalyzed by Rubisco it is still unknown what determines the efficiency with which carbon is fixed. Since all Rubisco enzymes examined exhibit both carboxylase and oxygenase activities, it has been argued that the oxygenase activity is an inevitable consequence of the active site chemistry (reviewed by Ellis & Gatenby, 1984); however the balance between the carboxylation and oxygenation has been shown to vary. The ratio of carboxylation to oxygenation has been defined by Laing *et al.* (1974) as follows:-

$$v_c/v_o = (V_c K_o / V_o K_c) ([CO_2]/[O_2])$$

where  $v_c$  and  $v_o$  are the velocities of carboxylation and oxygenation respectively,  $V_c$  and  $V_o$  the maximal velocities of the two reactions, and  $K_c$  and  $K_o$  the Michaelis constants for  $CO_2$  and  $O_2$ . The substrate specificity ( $V_c K_o / V_o K_c$ ) determines the rates of the two reactions at any given  $CO_2$  and  $O_2$  concentration; a high specificity value indicate a greater specificity for  $CO_2$ . Jordan & Ogren (1981; 1983) determined the substrate specificity factor for a variety of organisms and observed a low value for the form II Rubisco from *R.rubrum* (15), whereas the substrate specificity factor for the form I Rubisco from cyanobacteria and higher plants was approximately 48 and 80 respectively. Jordan and Ogren concluded that modifications increasing the carboxylation efficiency had occurred during evolution. The observation that the enzyme activity is lost from the hexadecameric Rubisco enzyme upon removal of the small subunits indicates their absolute requirement for catalysis.

### 1.2.3 SUBUNIT SYNTHESIS

In eukaryotes the Rubisco enzyme is found in the stromal compartment of the chloroplast. Since the first report in 1962 that chloroplasts contain all the components

required to carry out protein synthesis (reviewed by Kirk & Tilney-Bassett, 1978), much research has been carried out to determine which proteins are synthesized in this plant organelle (reviewed by Ellis, 1981). The use of specific inhibitors such as chloramphenicol and cycloheximide, which inhibit chloroplast 70S and cytosolic 80S ribosomes respectively, allowed the uncoupling of the two protein-synthesizing systems (Criddle *et al.*, 1970). The inhibition of chloroplast protein synthesis by chloramphenicol was shown to result in an inhibition of the synthesis of Rubisco large subunits in greening barley shoots, whereas the synthesis of the Rubisco small subunits was inhibited by cycloheximide. These observations led Criddle *et al.* (1970) to conclude that the Rubisco large subunits are synthesized inside the chloroplast while the small subunits are synthesized by cytosolic ribosomes. Uncoupling of the subunits using specific inhibitors has also been reported for *Pisum sativum* shoots (Cashmore, 1976), *Nicotiana tabacum* protoplasts (Hirai & Wildman, 1977) and *Glycine max* cells (Barraclough & Ellis, 1979).

In young leaves of several cereal seedlings such as *Secale cereale* L (rye), the formation of 70S but not 80S ribosomes is inhibited at an elevated temperature between 32 °C and 34 °C (Feierabend & Schrader-Reichhardt, 1976). This phenomenon was utilized to demonstrate that plants grown at the non-permissive temperature fail to synthesize Rubisco large subunits, but do synthesize Rubisco small subunits (Feierabend & Wildner, 1978). The demonstration that the large subunit is inherited in a non-Mendelian fashion in *Nicotiana tabacum* (Sakano *et al.*, 1974) and in *Chlamydomonas reinhardtii* (Spreitzer & Mets, 1980) provided genetic evidence that the Rubisco large subunit is not nuclear-encoded, and thus is probably encoded in the chloroplast genome.

Direct evidence that the Rubisco large subunits are synthesized on chloroplast ribosomes came from experiments with isolated intact chloroplasts. Blair & Ellis (1973) were the first to demonstrate that in intact chloroplast isolated from *Pisum sativum* leaves, protein synthesis can be driven by light energy if chloroplasts are incubated in an appropriate medium containing radiolabelled amino acids. The major polypeptide synthesized inside these isolated chloroplasts was identified as the large subunit of Rubisco. Synthesis of the large subunit has since been demonstrated in chloroplasts isolated from *Spinacia oleracea*,

*Euglena gracilis*, *Sorghum vulgare* and *Acetabularia* (reviewed by Ellis, 1981). Whereas these above observations provide good evidence that the large subunit is synthesized inside the chloroplast they only suggest that the large subunit may be encoded by the chloroplast genome. Direct evidence for this idea was obtained when chloroplast DNA, rather than RNA, was isolated from *Zea mays* leaves, transcribed *in vitro* by *E.coli* RNA polymerase, and the transcripts subsequently translated in a cell-free rabbit reticulocyte lysate. Among the large number of polypeptides synthesized the Rubisco large subunit was identified as the most abundant (Coen *et al.*, 1977). From these initial observations it was concluded that the Rubisco large subunit is encoded by the chloroplast genome and synthesized within the chloroplast.

A precursor of the large subunit with an apparent Mr of approximately 2000 larger than the mature polypeptide was identified when total RNA from *Spinacia oleracea* chloroplasts was translated in an *E.coli* cell-free system (Landridge, 1981). Treatment of the precursor with a soluble chloroplast extract converts it to its mature size indicating the presence of a processing activity. In support of possible post-translational processing of the large subunit is the observed discrepancy between the determined amino acid sequence of the mature large subunit from *Hordeum vulgare* (Poulsen, 1981), *Triticum aestivum* and *Sorghum vulgare* (Miziorko & Lorimer, 1983) and the deduced amino acid sequence (McIntosh *et al.*, 1980; Zurawski *et al.*, 1981; Dron *et al.*, 1982; Shinozaki & Sugiura, 1982), where the latter contains 14 additional amino acid residues at the aminoterminal end. Whether the processing occurs *in vivo* or *in vitro* during protein purification remains to be established. It is noted that in *Nicotiana tabacum* the determined (Amiri *et al.*, 1984) and the deduced (Shinozaki & Sugiura, 1982) amino acid sequence of the large subunit are identical at the aminoterminal end.

The Rubisco small subunit is a product of cytosolic protein synthesis since its synthesis is inhibited *in vivo* by cycloheximide but not by chloramphenicol (Criddle *et al.*, 1970; Cashmore, 1976; Barraclough & Ellis, 1979). Genetic analysis in *Nicotiana tabacum* showed a Mendelian inheritance for the Rubisco small subunit (Kawashima & Wildman, 1972), provided the first evidence that the small subunit is nuclear-encoded. Furthermore,



the small subunit mRNA was shown to be translated *in vitro* on free cytosolic ribosomes to give a precursor polypeptide with an apparent Mr about 6000 larger than the mature polypeptide (Dobberstein *et al.*, 1977; Cashmore *et al.*, 1978; Highfield & Ellis, 1978), whereas transcripts coding for the small subunit precursor have been identified in both the nuclear and cytosolic compartments of leaf cells from *Pisum sativum* (Smith & Ellis, 1981; Gallagher & Ellis, 1982).

The synthesis of Rubisco in prokaryotes is less complex due to the lack of cellular compartments. In cyanobacteria the small subunit is synthesized as the mature polypeptide, and there is no evidence for the synthesis of a large subunit precursor.

The observation that the synthesis of the Rubisco subunits in eukaryotes occurs in two different cellular compartments, whereas the assembled Rubisco enzyme is found only inside the chloroplast implies that the small subunit precursor is transported across the chloroplast envelope prior to assembly (see Section 1.2.4) and suggests that a coordinate regulation of Rubisco subunit synthesis may exist (see Section 1.2.5).

#### 1.2.4 SUBUNIT ASSEMBLY

Assembly of the Rubisco enzyme has been studied in a variety of different experimental systems, including assembly in whole leaves (Roy *et al.*, 1978; 1979), intact isolated chloroplasts (Blair & Ellis, 1973; Barraclough & Ellis, 1980) and chloroplast extracts (Milos & Roy, 1984). Reconstitution of Rubisco assembly has been studied in isolated yeast mitochondria (Hurt *et al.*, 1986), in *E. coli* (reviewed in Section 1.4) and *in vitro* (Voordouw *et al.*, 1984). This section discusses a number of results obtained from these systems which have contributed to the current model of Rubisco assembly as presented in Figure 4.

The Rubisco small subunit is synthesized as a precursor by the cytosolic ribosomes in eukaryotic cells (see Section 1.2.3) and transported across the two envelope membranes. This transport occurs post-translationally since it has been shown that subunit precursors are imported after synthesis into intact chloroplasts isolated from *Spinacia oleracea* (Chua &

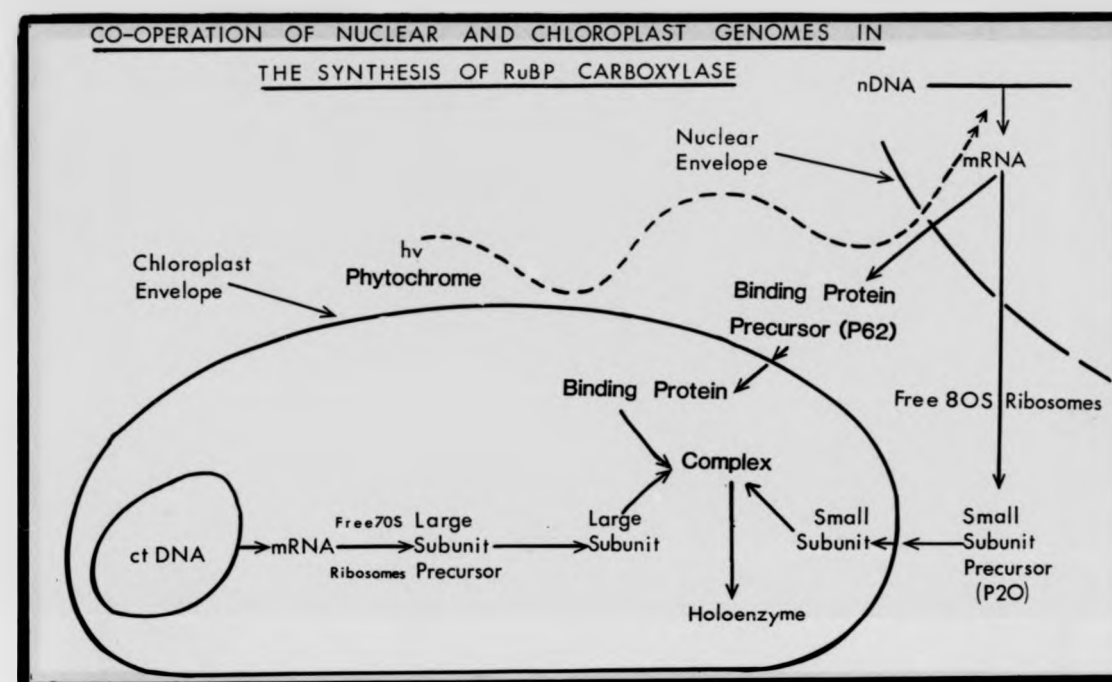


Figure 4. Synthesis and assembly of hexadecameric Rubisco

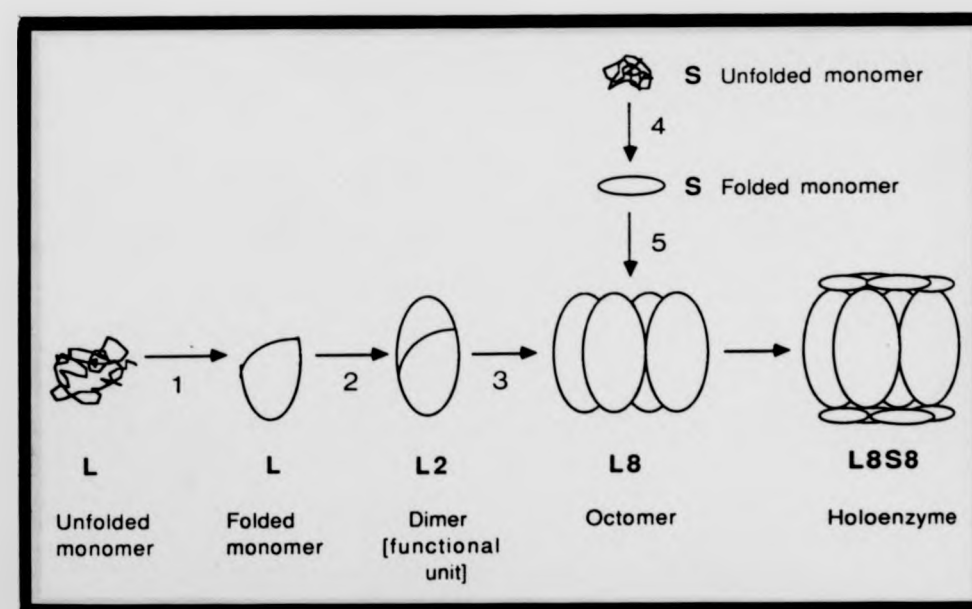
Panel A shows a suggested pathway for the synthesis and assembly of Rubisco from plants. Rubisco large subunits are synthesized from chloroplast genes (ctDNA) whereas Rubisco small subunits and the binding protein subunits are synthesized from nuclear genes (nDNA). The symbols P20 and P62 refer to the apparent molecular masses  $\times 10^{-3}$  for the precursor polypeptides respectively. The wavy line indicates the role of light in the stimulation of the small subunit gene transcription. This diagram was taken from a paper by Ellis (1987).

Panel B shows the proposed order of events in the formation of the hexadecameric Rubisco enzyme: 1. folding of the large subunit (L) monomer; 2. formation of the large subunit dimer (L2); 3. tertamerization of large subunit dimers to give the octameric (L8) core; 4. folding of the small subunit (S) monomer; 5. association of the small subunits with the L8 core to give the hexadecameric holoenzyme. This diagram was adapted from a paper by Goloubinoff *et al.* (1989)

A



B



Schmidt, 1978) and from *Pisum sativum* (Highfield & Ellis, 1978). These observations are in agreement with the failure to find cytosolic ribosomes bound to the chloroplast envelope during chloroplast development (Ellis, 1981). This type of protein transport thus differs from the observed co-translational translocation of secretory proteins across the membrane of the endoplasmic reticulum, which is dependent on concomitant protein synthesis (reviewed by Walter *et al.*, 1984)

The small subunit precursors bind to the outer surface of isolated intact chloroplasts. When the chloroplasts are pretreated with proteases the subsequent uptake of small subunit precursors is inhibited (Chua & Schmidt, 1978; Cline *et al.*, 1985), suggesting that membrane receptor proteins probably interact with the precursor polypeptides. Precursors which are bound to isolated envelope membranes were shown to be sensitive to low concentrations of trypsin (Pfisterer *et al.*, 1982). The sensitivity of proteins to proteases is often used as a measure for the folded state of the protein and the protease-sensitivity of the bound precursor molecules thus suggesting a partially folded conformation.

Grossman *et al.* (1980) showed that energy in the form of ATP is required to move the precursors across the chloroplast membranes. Either during or after transport the small subunit precursor is processed to its mature size by a highly specific protease located in the stromal fraction (Dobberstein *et al.*, 1977; Smith & Ellis, 1979). The processing of the 20 kDa small subunit precursor from *Pisum sativum* to the 14 kDa mature polypeptide is independent of ATP and proceeds in two steps. Robinson & Ellis (1984) observed a 18 kDa intermediate form when small subunit precursors were treated with the partially purified stromal protease. The significance of this two-step processing is unknown. It has recently been shown that import of the small subunit precursor into isolated intact chloroplast requires a chloroplast ATPase (Pain & Blobel, 1987). The function of this ATPase is unknown at present since it does not generate a membrane potential which could be utilized to drive import. The small subunit presequence which is thought to contain the information required for transport into chloroplast, has successfully been used to target foreign bacterial proteins into chloroplasts both *in vivo* and *in vitro* (recently reviewed by Keegstra, 1989).

Although the synthesis of the Rubisco subunits and the transport of the small subunit precursor into chloroplasts are well-studied processes (Fig. 4A), the mechanism by which the two types of subunits assemble into a hexadecameric protein complex which is biologically functional is poorly understood. The available evidence suggests a complex pathway for Rubisco assembly. The observations that newly-synthesized Rubisco large subunits assemble into holoenzyme molecules in isolated intact chloroplasts (Barracough & Ellis, 1980) suggests the existence of a pool of assembly-competent free small subunits. The possibility that the newly-synthesized large subunits either assemble with small subunits that have been released from pre-existing Rubisco enzyme complexes or exchange with large subunits within these complexes is unlikely but cannot be ruled out. Pools of free cytosolically synthesized small subunits have also been identified in extracts of leaves of *Pisum sativum* (Roy *et al.*, 1978; 1979). Unassembled small subunits are rapidly degraded in *Chlamydomonas reinhardtii* (Schmidt & Mishkind, 1983), and have a half-life of less than 7.5 min, which suggests that the pool of unassembled subunits is probably small, but sufficient to permit Rubisco assembly for a short time *in vitro*.

Barracough & Ellis (1980) were the first to show that in intact chloroplasts isolated from *Pisum sativum*, the newly-synthesized large subunits bind to another abundant stromal protein which they termed the Rubisco large subunit binding protein. Analysis of the newly-synthesized large subunits from either *P. sativum* seedlings or isolated intact chloroplasts on sucrose density gradients showed that they sediment as two distinct forms with sedimentation values quoted as 7S and 29S (Roy *et al.*, 1982). The 7S complex has been suggested to represent large subunit dimers, while the 29S complex is identical to the Rubisco large subunit binding protein complex identified by Barracough and Ellis (1980). The binding of the large subunits to this binding protein also occurs *in vivo* (Ellis, 1981), and time course experiments with isolated intact chloroplasts indicated that the large subunits bind to this protein prior to assembly into the holoenzyme. Barracough & Ellis (1980) proposed that this association may be an intermediate step in the pathway of Rubisco assembly. The properties of the Rubisco large subunit binding protein and its possible role in assembly of the Rubisco holoenzyme are reviewed in Section 1.3.

The demonstration that bacterial proteins can be imported into isolated intact chloroplasts providing that they contain the Rubisco small subunit presequence at the aminotermminus (reviewed by Cashmore *et al.*, 1985), caused Hurt *et al.* (1986) to use a similar approach to investigate import of chloroplast protein into intact mitochondria isolated from *Saccharomyces cerevisiae*. When the presequence of cytochrome *c* oxidase subunit IV (a protein normally transported into mitochondria) was fused to the aminotermminus of the Rubisco large and mature small subunit from *Chlamydomonas reinhardtii*, these fusion proteins were efficiently imported into isolated intact yeast mitochondria. In mitochondria containing both processed Rubisco large and small subunits, no formation of oligomeric protein complexes was observed as judged by migration on sucrose density gradients (Hurt *et al.*, 1986). Since mitochondria normally do not contain Rubisco the mitochondrial matrix may not be a suitable environment for assembly, and the failure to assemble the Rubisco subunits may indicate the requirement for the chloroplast Rubisco large subunit binding protein. This idea is supported by the recent observation that cyanobacterial Rubisco large subunits assemble into holoenzyme oligomers after being imported as fusion proteins containing the soybean Rubisco small subunit presequence into intact chloroplasts isolated from *P. sativum* (Gatenby *et al.*, 1988).

A traditional means to investigate the folding and assembly of proteins is to study the dissociation and reassociation of purified protein *in vitro*. Treatment of the hexadecameric Rubisco purified from *Spinacia oleracea* with a low concentration of urea (3-4M) causes the small subunits to unfold and dissociate from the oligomeric protein complex with proportional loss of enzyme activity (Voordouw *et al.*, 1984). The sedimentation coefficient of the remaining protein structure ( $s_{20,w}=15.0$  S) indicates a conformation close to an octamer, possibly with reduced stability. The reconstitution of an enzymically active enzyme by removing the urea either in the presence or the absence of the small subunits has been unsuccessful because the large subunits aggregate and form insoluble complexes (Voordouw *et al.*, 1984). The Rubisco purified from the cyanobacteria *Synechococcus* and *Prochloron* is a hexadecameric enzyme from which small subunits can be isolated in monomeric form after incubation at pH 5.3 (Andrews & Ballment, 1983; Andrews *et al.*, 1984). The remaining

large subunits are stable octamers which retained the ability to assemble with either homologous or heterologous small subunits into functional hexadecamers (Andrews *et al.*, 1984; Andrews & Lorimer, 1985; Incharoensakdi *et al.*, 1985). Assembly of a hybrid Rubisco from its subunits synthesized in *E.coli* has also been reported (see Section 1.4.3.; van der Vies *et al.*, 1986).

Since the dissociation of the small subunits from the *S.oleracea* holoenzyme leaves the large subunits in a loose and flexible octameric form (Voordouw *et al.*, 1984), this may allow inter and/or intra specific interactions to occur more frequently than in the assembled holoenzyme. Some of these interactions may lead to new conformations of the large subunits that are no longer capable of binding small subunits and are no longer soluble in aqueous solution. Thus it seems likely that reduction of the flexibility of the *S.oleracea* large subunit octamer will decrease the frequency of inter and/or intra specific interactions and hence may prevent aggregates to be formed. This idea is supported by the recent observation that when Rubisco purified from *N.tabacum* is immobilized on Sepharose 4B beads the loss of enzymic activity caused by the urea-dependent dissociation of the small subunits is fully restored when the concentration of urea is reduced to 0.5M in the presence of isolated small subunits (Liren *et al.*, 1988). These observations indicate the requirement for a factor which stabilizes the large subunit octamers. The stabilization of the large subunits and/or assistance in the assembly process *in vivo*, is a function which has been proposed for the Rubisco large subunit binding protein (Barraclough & Ellis, 1980).

#### 1.2.5. GENE ORGANISATION AND EXPRESSION

The genes for the Rubisco large and small subunits in plants are located in the chloroplast and nuclear genomes respectively (Section 1.2.3). The gene encoding the large subunit is present as a single copy per chloroplast DNA molecule (Bedbrook *et al.*, 1979). However, each chloroplast contains 10 to 100 copies of the DNA molecule, and each leaf cell may contain 10 to 200 chloroplasts depending on the species and the developmental state of the cell (Lamppa & Bendich, 1979; Mizioro & Lorimer, 1983). It thus follows that there

may be several thousand copies of the large subunit gene per cell. Since the isolation and sequence determination of the large subunit gene from *Zea mays* (McIntosh *et al.*, 1980), corresponding genes from a variety of other plant species have been isolated and characterized (reviewed by Rochaix, 1985). Although no introns have yet been found in the large subunit genes of plants, nine introns have been identified in the large subunit gene from *Euglena gracilis* (Koller *et al.*, 1984). The arrangement of the Rubisco large subunit gene in the chloroplast genomes of plants is remarkably conserved, even among the distantly related monocotyledons and dicotyledons (Rochaix, 1985).

In plants, the Rubisco small subunit polypeptide is encoded by a small multigene family (Berry-Lowe *et al.*, 1982; Broglie *et al.*, 1983; Cashmore, 1983; Coruzzi *et al.*, 1983). The small subunit genes have been shown to contain introns and their number depends on the plant species. In monocots the small subunit gene contains one intron; in most dicots the genes contain two introns (reviewed by Nagy *et al.*, 1986), whereas a third intron has been identified in *Nicotiana tabacum* (Mazur & Chui, 1985). The introns in the monocot and dicot small subunit genes are situated at identical positions with respect to the coding sequence and this suggests that introns may have been lost in evolution. Interestingly the first exon encodes the entire presequence and either one or two amino acids of the mature polypeptide, a finding which supports Gilbert's (1979) proposal that exons encode functional polypeptide domains (Mizioko & Lorimer, 1983). The members of the small subunit gene family have been shown to be linked in the nuclear genome either in tandem (Dean *et al.*, 1985) or in a reversed orientation (Cashmore, 1983), suggesting that they may have arisen by gene duplication.

In the prokaryotic cyanobacteria the Rubisco large and small subunit genes are linked and cotranscribed from a contiguous piece of DNA (Shinozaki & Sugiura, 1983; Nierzwicki-Bauer *et al.*, 1984). The eukaryotic rodophytic (chlorophyll *a*, phycobilin-containing) algae represent ancient evolutionary lineages which are distinct from the chlorophyllic (chlorophyll *a,b*-containing) plants (reviewed by Cattolica, 1986). In those algae the cyanelle genome contains a single Rubisco operon that contains both the large and small subunit genes (Heinhorst & Shively, 1983; Steinmuller *et al.*, 1983; Reith & Cattolica, 1986). It is attractive

to speculate that these photosynthetic cyanelles represent an intermediate stage between prokaryotic endosymbionts and photosynthetic cell organelles.

The separate sites of synthesis of the Rubisco large and small subunits in plants, and the imbalance in the copy number of their genes, has implications for the regulatory mechanisms that may ensure the required production of subunits for Rubisco assembly. This regulation can occur either at the level of transcription, or RNA splicing or translation, but can also be controlled by the stability of transcripts and polypeptides. From the current wealth of available information it is clear that in many plant species this regulation occurs at a number of different steps and is often dependent on light, temperature and the developmental stage of a particular tissue. It is outside the scope of this section to review this area of research in detail; in contrast a selected number of observations are discussed to illustrate the current understanding of Rubisco gene expression and regulation.

When dark-grown *Pisum sativum* seedlings are illuminated the amount of Rubisco protein accumulates (Smith & Ellis, 1981) mainly through an increase in the abundance of both the large and small subunit mRNAs (Gallagher *et al.*, 1984). This light-dependent accumulation is not an essential requirement for Rubisco synthesis, since the enzyme is also detected in dark-grown cereal leaves (Ellis, 1978) and the light-dependent accumulation varies between different plant species (reviewed by Tobin & Silverthorne, 1985). In *Zea mays* the amounts of large subunit mRNA in dark-grown and light-grown plants is similar. However the presence of two different sizes of large subunit transcript, that alter in relative abundance during light-induced plastid development, may indicate an alternative mechanism to control Rubisco large subunit accumulation.

The transcription of the small subunit genes proceeds at a faster rate in nuclei isolated from light-grown *Pisum sativum* leaves than in nuclei isolated from dark-grown apical buds and this observation led Gallagher *et al.* (1984) to conclude that the control of small subunit synthesis operates at either the rate of transcription or the rate of RNA turnover. The latter possibility is unlikely to be correct since the small subunit transcripts are stable in isolated dark-grown nuclei at least for 3 hours. An example of Rubisco synthesis regulated at the level of translation is found in *Amaranthus hyochondriacus* (Berry *et al.*, 1986). The



illumination of dark-grown cotyledons rapidly increases the synthesis of both the large and small subunits before any increase in the levels of mRNA is observed. Recently Klein & Mullet (1987) showed that both large subunit synthesis and the level of large subunit mRNA decrease when *Hordeum vulgare* seedlings are illuminated. Detailed studies on light-regulation have revealed that in *Lemna gibba* and *Pisum sativum* the level of small subunit transcripts in etiolated plants can be evaluated by a pulse of red light; this induction can be reversed by far-red light, suggesting that the effect is mediated by the photomorphogenic pigment phytochrome (reviewed by Ellis, 1986; Nagy *et al.*, 1986). This phytochrome protein contains the chromophore tetrapyrrole and exists in two forms i.e. biologically active and non-active which are interconvertable by radiation. The mode of action of phytochrome is unknown.

The expression of different members of the small subunit gene family in *Petunia hybrida* has been investigated by Dean *et al.* (1985) who showed that, of the eight small subunit genes present, two show no detectable expression, five give rise to between 2% and 23% of the total small subunit transcripts, while one accounts for 47% of the total small subunit transcripts in green leaves. Similar observations have been made in *Pisum sativum* where one member of the five-members-counting family is responsible for about 35% of the total small subunit transcripts in green leaves (Coruzzi *et al.*, 1984). The biological significance of this differential expression is unknown, but it may serve a developmental purpose.

Since the small subunit genes encode a subunit of a photosynthetic enzyme it is perhaps not surprisingly that in most plants, these genes are expressed only in organs that contain chloroplasts. Although the small subunit mRNA is a major transcript in green leaves from *Pisum sativum*, other organs such as petals, pods and seeds also contain a substantial amount of the small subunit mRNA. Coruzzi *et al.* (1984) showed that, depending on their ages, stems contain 50% of the total small subunit mRNA level in leaves, whereas roots contain less than 1% of the small subunit mRNA level found in leaves. The different levels of small subunit mRNA in the various organs may well be a reflection of variable amounts of chloroplast-containing cells in some of these organs (Eckens *et al.*, 1985). In leaves of C4



plants, such as *Zea mays*, both the large and small subunit polypeptides are present in the bundle sheath cells but not in the mesophyll cells. This differential expression is limited to the light-stimulated accumulation of both the large and small subunit transcripts (Link *et al.*, 1978; Broglie, Coruzzi, Keith & Chua, 1984).

The ability to transfer DNA into plant nuclear chromosomes using *Agrobacterium tumefaciens* Ti-based plasmids has allowed the analysis of sequences responsible for the light-dependent regulation of the Rubisco small subunit genes. When the small subunit structural gene from *Pisum sativum* plus 1 kb of 5' flanking sequence upstream of the transcriptional start point, is transferred to the nuclear genome of either *Petunia hybrida* (Broglie, Coruzzi, Fraley, Rogers, Horsch, Niedermeyer, Fink, Flick & Chua, 1984) or *Nicotiana tabacum* (Herrera-Estrella *et al.*, 1984) the light-dependent regulation of the small subunit gene is similar to that observed in pea leaves. A region in the 5' flanking DNA sequence on which light regulation depended was subsequently identified. When this region was fused to the bacterial gene encoding chloramphenicol acetyl transferase, and the fused gene transferred to the nuclear genome of *N. tabacum*, regulation of the bacterial protein was shown to be identical to the regulation of the small subunit gene in its normal situation (Herrera-Estrella *et al.*, 1984). This regulatory region functions independently of its orientation and thus behaves as an enhancer-like element (Timko *et al.*, 1985)

### 1.3 RUBISCO LARGE SUBUNIT BINDING PROTEIN

#### 1.3.1 DISCOVERY

The Rubisco large subunit binding protein was discovered by Barraclough & Ellis (1980) during studies on the synthesis of proteins in intact chloroplasts isolated from *Pisum sativum*. When intact isolated chloroplast are incubated in a suitable medium containing K<sup>+</sup> ions, <sup>35</sup>S-methionine and light as an energy source, they synthesize a number of labelled polypeptides, among which the most abundant is the Rubisco large subunit (Blair & Ellis, 1973). Analysis of the physical state of these newly-synthesized large subunits under non-

denaturing conditions showed that a minority of these subunits are assembled in the holoenzyme. The synthesis of the large subunits proceeds at a faster rate than their assembly and Barraclough & Ellis (1980) observed that the newly-synthesized unassembled large subunits comigrate on non-denaturing polyacrylamide gels with another abundant stromal protein of an apparent Mr > 750 000. This comigration was shown to be exact as it was observed over a range of polyacrylamide concentrations, indicating that the two types of polypeptides are bound together. Barraclough and Ellis named this protein the Rubisco large subunit binding protein (binding protein) and proposed that the binding of the large subunits to this protein may be required for the assembly of the Rubisco enzyme. Roy *et al.* (1982) analyzed radiolabelled products of both isolated chloroplasts and whole leaves of *P. sativum* on sucrose density gradients, and showed that a pool of Rubisco large subunits is associated with the binding protein confirming the earlier observations made by Barraclough & Ellis (1980). It has been proposed that the binding protein keeps the Rubisco large subunits in a form suitable for assembly with the small subunits to produce the enzymically active oligomer (Ellis, 1981).

The evidence for the involvement of the Rubisco large subunit binding protein in the assembly of the Rubisco holoenzyme is still circumstantial. The next section reviews the physical and biochemical properties of the Rubisco large subunit binding protein that were known when the current work commenced. The general discussion surveys the position as it was when this thesis was completed.

### 1.3.2 PURIFICATION

The Rubisco large subunit binding protein has been purified from *Pisum sativum* (Hemmingsen & Ellis, 1986) and *Hordeum vulgare* chloroplasts (Johnson, 1987; Musgrove *et al.*, 1987) as a protein with an apparent Mr of approximately 720 000 as determined by gelfiltration chromatography. Analysis of the binding protein purified from *P. sativum* on polyacrylamide gels containing SDS revealed the presence of two polypeptides with apparent Mrs of approximately 61 000 (termed the  $\alpha$  subunit) and 60 000 (termed the  $\beta$  subunit)

which are present in equal amounts in the purified oligomer. The binding protein from *H.vulgare* and *Triticum aestivum* consists of  $\alpha$  and  $\beta$  subunits with a slightly higher Mr than those from *P.sativum* (Musgrove *et al.*, 1987; see Section 3.1.2). The two subunits from *P.sativum* are immunologically distinct and show different partial protease digestion patterns (Musgrove *et al.*, 1987). Comparison of their determined aminoterminal sequences shows that the  $\alpha$  and  $\beta$  subunits are different but probably related, since optimized alignment of the two sequences shows 26% identical amino acids (Ellis *et al.*, 1987),

Recently a protein with an apparent Mr of about 700 000 has been purified from the photosynthetic prokaryote *Chromotium vinosum* and was shown to be immunologically related to the Rubisco large subunit binding protein (Torres-Ruiz & McFadden, 1988). This protein is a homo-oligomer of polypeptides with an apparent Mr of approximately 60 000. The determined aminoterminal sequence of this bacterial polypeptide shows 38% amino acid identity with both the Rubisco large subunit binding protein  $\alpha$  and  $\beta$  subunit aminoterminal sequences from *P.sativum*. Whether this bacterial homologous protein is involved in Rubisco assembly in *C.vinosum* remains to be determined.

### 1.3.3 SYNTHESIS AND OCCURRENCE

The first indication that the Rubisco large subunit binding protein subunits are synthesized on cytosolic ribosomes came from the observation that the binding protein is not labelled when intact chloroplasts isolated from *Pisum sativum* were incubated with radiolabelled amino acids (Blair & Ellis, 1973). More recently it has been shown that when the mRNAs present on the cytosolic polysomes isolated from *P.sativum* leaves are translated *in vitro* using a wheatgerm cell-free translation system in the presence of  $^{35}\text{S}$ -methionine and chloramphenicol, a  $^{35}\text{S}$ -methionine-labelled protein is precipitable using antibodies raised against the Rubisco large subunit binding protein oligomer. The Mr of this immunoprecipitated protein is slightly higher than that of the mature  $\alpha$  and  $\beta$  binding protein subunits as judged by SDS polyacrylamide gelelectrophoresis (Hemmingsen & Ellis, 1986). The synthesis of a polypeptide with a higher molecular mass suggests that the chloroplast

binding protein is encoded by nuclear genes and its subunits synthesized in the cytosol as precursors. The precursors to the  $\alpha$  and  $\beta$  subunits are thought to have very similar apparent  $M_r$ s as the above mentioned immunoprecipitated polypeptide is not resolved into two bands on polyacrylamide gels under conditions that allow separation of the mature binding protein subunits (Hemmingsen & Ellis, 1986).

Lennox & Ellis (1986) compared the amounts of Rubisco and Rubisco large subunit binding protein in *P. sativum* plants grown under different light regimes; rocket immuno electrophoresis was used to ensure absolute quantitation of the two proteins. The main increase in the amount of Rubisco was observed 24 h after 8 days-old etiolated pea plants were exposed to continuous white light. The main increase in the amount of Rubisco large subunit binding protein was also observed after 24 h of illumination; however the increase of the binding protein is lower than that observed for the Rubisco enzyme. Dark-grown pea plants contain detectable amounts of both these proteins; the binding protein and the Rubisco enzyme account for 1.8% and 14% of the total soluble protein respectively. In plants grown for the same period under a photoperiod of 12 h the binding protein comprises 2.4% of the total soluble protein whereas the amount of Rubisco increases to 83% (Lennox & Ellis, 1986). Thus the light-induced accumulation of the binding protein is much lower than the accumulation of the Rubisco enzyme which is consistent with a possible catalytic role of the binding protein in the assembly of the Rubisco oligomer.

Antibodies raised against the *Pisum sativum* Rubisco large subunit binding protein oligomer recognize polypeptides with apparent  $M_r$ s of about 60 000 in extracts of leaves of various plants such as *Spinacia oleracea*, *Nicotiana tabacum*, *Triticum aestivum* and *Hordeum vulgare*, as well as in plastids isolated from the endosperm of both developing and germinating *Ricinus communis* seeds (Hemmingsen & Ellis, 1986). These colourless, non-photosynthetic plastids, called leucoplasts, contain the Rubisco enzyme for unknown purposes (Benedict, 1973), and are able to import and process Rubisco small subunit precursor polypeptides *in vitro* (Boyle *et al.*, 1986). In the C4 plant *Zea mays*, the binding protein is present in the bundle sheath cells that contain Rubisco, whereas it is not found in the mesophyll cells that do not contain Rubisco (Ellis *et al.*, 1987). The abundance of the

binding protein is either very low or below the detection limit in plant tissues which lack Rubisco such as roots (S.M. Hemmingsen, personal communication). These observations indicate a co-distribution of the binding protein and the Rubisco enzyme and are consistent with the hypothesis that the binding protein is involved in Rubisco assembly. Polypeptides immunologically related to the Rubisco large subunit binding protein have also been detected in extracts of bacteria, some of which do not contain Rubisco, such as *Escherichia coli* (S.M. Hemmingsen, personal communication; see also Section 3.3.2).

#### 1.3.4 DISSOCIATION BY MgATP

The addition of MgATP to extracts of *Pisum sativum* chloroplasts causes the dissociation of the Rubisco large subunit binding protein oligomer (Bloom *et al.*, 1983; Hemmingsen & Ellis, 1986; Musgrove & Ellis, 1986; Musgrove *et al.*, 1987). This dissociation by MgATP is highly specific since other nucleotides such as GTP, ADP, CTP, AMP, NADH, 3'-5'cyclic AMP, pyrophosphate and the non-hydrolysable  $\beta$ - $\gamma$ -x-methylene analogue of ATP in the presence of  $Mg^{2+}$  ions do not cause dissociation (Musgrove & Ellis, 1986). It has been suggested that the hydrolysis of ATP may be necessary for the dissociation of the binding protein oligomer (Bloom *et al.*, 1983). However the effect of MgATP is not associated with either the phosphorylation or the adenylation of the binding protein subunits in stromal extracts (Hemmingsen & Ellis, 1986). Instead the purified Rubisco large subunit binding protein from *P. sativum* has been reported to exhibit low levels of ATPase activity (Candhari *et al.*, 1987).

Hemmingsen & Ellis (1986) showed that the dissociation of the binding protein by MgATP is reversible when the ATP concentration in stromal extracts is lowered by allowing protein synthesis to proceed. However when MgATP is removed by dialysis the reversible dissociation is not observed suggesting that dialysable factors may be required for this reversal (Musgrove & Ellis, 1986). The detection of a low amount of binding protein monomeric subunits in the absence of MgATP, and the presence of some binding protein oligomer at a concentration of 5 mM MgATP, led Lennox and Ellis (1986) to propose that

the binding protein oligomeric and monomeric forms are in equilibrium, and that MgATP causes this equilibrium to shift towards the monomeric form. In the presence of 0.5 mM MgATP, dissociation of the binding protein is observed *in vitro*. Under normal physiological conditions the ATP concentration in the chloroplast *in vivo* has been calculated to be 1-3mM (Hampp *et al.*, 1982) and this concentration would favour dissociation. However the concentration of the binding protein in stromal extracts is approximately 20 to 100 fold lower than the binding protein concentration inside the chloroplast which is estimated to be about 10 mg/ml (Lennox & Ellis, 1986). Increasing the concentration of the binding protein will strongly favour association of the subunits, thus the position of the equilibrium *in vivo* remains to be determined. The effect of MgATP on the binding protein oligomer-monomer equilibrium was shown to be less extensive in chloroplast extracts which had been concentrated 15-fold (Roy, Hubbs & Cannon, 1988). Whether this observation is either the result of an increased protein concentration or the result of the loss of molecules with an apparent Mr < 50 000 from the stromal fraction, due to the concentration procedure used, is not known.

The fate of the associated newly-synthesized Rubisco large subunits upon dissociation of the binding protein oligomer in the presence of MgATP in stromal extracts is unclear. Hemmingsen & Ellis (1986) observed that the Rubisco large subunits migrate near the top of the sucrose density gradient when stromal fractions of *P.sativum* are treated with MgATP. The sucrose gradient fractions in which the Rubisco large subunits were detected also contain binding protein polypeptides, but whether these binding protein polypeptides form heterodimers with the Rubisco large subunits is unknown. Roy, Cannon & Gilson (1988) reported that when stromal extracts of *P.sativum* are incubated with MgATP the Rubisco large subunits leave the binding protein and comigrate with the Rubisco enzyme (Milos & Roy, 1984; Milos *et al.*, 1985). This transfer of labelled large subunits to the holoenzyme is inhibited by antibodies against the purified Rubisco large subunit binding protein (Cannon *et al.*, 1986). The transfer of the labelled large subunits may suggest their assembly into the Rubisco holoenzyme; however it could also indicate either an exchange and/or association of the labelled large subunits with already assembled large subunits in the

holoenzyme. Thus the above observations are in agreement with the hypothesis that the binding protein is involved in Rubisco assembly, but it remains to be determined whether the association of the large subunits with the binding protein is an obligatory step in the assembly pathway. This determination will require the development of an *in vitro* system in which the Rubisco oligomer can be produced from defined components.

### 1.3.5 THE RUBISCO LARGE SUBUNIT BINDING PROTEIN : - A MOLECULAR CHAPERONE ?

Very little is known about the mechanisms that are involved in the processes of folding and assembly *in vivo*, but folding and assembly are assumed to occur in an ordered fashion and are not the result of at random inter and/or intra specific interactions (Creighton, 1984). Some denatured proteins renature *in vitro* into functionally active conformations, however this is not generally true for all polypeptides e.g. Rubisco. The refolding of denatured proteins *in vitro* often only occurs under non-physiological conditions such as high pH, high ionic strength, different temperature and pressure (Creighton, 1984). Thus the failure of refolding and assembly *in vitro* may indicate either the requirement for non-physiological conditions or the need for a protein which function is to assist the folding/assembly process, a role which has been proposed for the Rubisco large subunit binding protein.

A precedent for such an assembly protein is nucleoplasmin which is required for the correct assembly of nucleosomes from DNA and histones in extracts of *Xenopus laevis* eggs (Laskey *et al.*, 1978). Nucleoplasmin does not bind to the nucleosomes or the DNA, but to the histones in a way that it shields their positive charges. Laskey and Earnshaw (1980) showed that nucleoplasmin promotes histone-histone interactions by competing with the electrostatic interactions between the histones and the DNA, thereby preventing the formation of insoluble complexes. The term 'molecular chaperone' was used by Laskey *et al.* (1978) to describe the role of nucleoplasmin which is required for the assembly of the nucleosome but is not a component of the final structure. This role is analogous to that

proposed for the Rubisco subunit binding protein in Rubisco assembly. Literature searches have revealed that the same term is appropriate to describe the role of certain other proteins, leading Ellis (1987) to propose that this term be extended to a wider range of proteins than nucleoplasmin.

A protein found in pre-B cells has been shown to bind non-covalently immunoglobulin heavy chains that have no light chains attached; hence this protein was termed the immunoglobulin heavy chain binding protein or BiP (Haas & Wahl, 1983). It has been suggested that BiP recognizes regions in the heavy chain that are covered by light chain upon assembly of the immunoglobulin molecule. These observations are strikingly similar to the observed association of the Rubisco large subunit and the binding protein in the chloroplasts of *P.sativum* and possibly suggest that both BiP and the chloroplast binding protein may be involved in controlling the conformation of one type of polypeptide, thereby allowing interaction with another type of polypeptide to form a functional oligomeric complex. BiP has been reported to bind to some specific aberrant proteins, such as mutants of influenza hemagglutinin and a fusion protein of the SV40 T antigen that are delivered to the endoplasmic reticulum (Gething *et al.*, 1986; Sharma *et al.*, 1985). The immunoglobulin heavy chains are released from BiP by MgATP, analogous to the observed release of the heat shock protein (hsp) 70 from heat-shocked nuclear structures. BiP and hsp 70 are also sequence-related, suggesting that there may be a family of 'binding proteins' (Bole *et al.*, 1986; Munro & Pelham, 1986). It has been postulated that this family of BiP-like proteins prevents aggregation of other proteins by binding to the hydrophobic surfaces in an ATP-reversible manner (Munro & Pelham, 1986). It was proposed by Ellis (1987) to term such proteins 'molecular chaperones' since their role may be to prevent certain protein interactions that occur during the assembly process which lead to the formation of non-functional complexes (Ellis *et al.*, 1987). Whether BiP and/or the Rubisco large subunit binding protein can be regarded as molecular chaperones remains to be determined. The number of proteins that are considered molecular chaperones is increasing rapidly and the different classes that are currently recognized within the molecular chaperone family are discussed in Section 4 (Table 21; Ellis, van der Vies & Alldrick, 1989)



## 1.4 SYNTHESIS AND ASSEMBLY OF RUBISCO SUBUNITS IN *ESCHERICHIA COLI*

### 1.4.1 SYNTHESIS OF PLANT RUBISCO SUBUNITS

The synthesis of the Rubisco large subunits from *Zea mays* and *Triticum aestivum* in *E.coli* was first reported by Gatenby *et al.* (1981). In these studies a chloroplast DNA fragment from either maize or wheat was used containing the large subunit structural gene plus sequences probably involved in gene expression and protein synthesis in chloroplasts. It was shown that *E.coli* cells harbouring a plasmid containing either the wheat or maize chloroplast DNA fragment, synthesize a polypeptide that is precipitable with antibodies raised against the Rubisco holoenzyme. This polypeptide has an apparent Mr of 55 000 as judged by SDS polyacrylamide gel electrophoresis, similar to the apparent Mr of the Rubisco large subunit isolated from chloroplasts. The synthesis of the large subunit polypeptide was independent of the orientation of the chloroplast DNA fragment in the plasmid vector, suggesting that the chloroplast transcription and translation signals are functional in *E.coli* (Gatenby & Cuellar, 1985; Bradley & Gatenby, 1985). Expression of the Rubisco large subunit gene in *E.coli*, using the endogenous chloroplast promoter, has also been reported for *Nicotiana tabacum* (Fluhr *et al.*, 1983; Zhu *et al.*, 1984), *Petunia hybrida* (Bovenberg *et al.*, 1984) and *Chlamydomonas reinhardtii* (Zhu *et al.*, 1984).

The amount of Rubisco large subunits synthesized in these various systems is low and does not allow detailed analysis. However an improved level of synthesis was obtained by increasing the rate of transcription of the large subunit gene. The amount of maize large subunit is increased to a maximum of 2% of the total *E.coli* protein when the large subunit is present on a multicopy plasmid and gene expression is controlled by the bacteriophage lambda P<sub>L</sub> promoter in the presence of a transcription termination protein (Gatenby & Castleton, 1982). A high level of synthesis is also achieved using the bacterial *lac* promoter, plus either the endogenous chloroplast ribosome binding site (Bradley *et al.*,

1986) or a hybrid ribosome binding site formed by fusing *lacZ* and the large subunit gene sequences (Somerville *et al.*, 1983; 1986).

Subsequent studies on the properties of the Rubisco large subunits in *E.coli* showed that the maize large subunits accumulate as electron-dense amorphous granules in the *E.coli* cell (Somerville *et al.*, 1986). When such *E.coli* cells are lysed the large subunits are detected in a stable but insoluble aggregated form. Even at a low level of synthesis, the maize large subunits are insoluble (Gatenby, 1984). Attempts to detect either enzymic activity or the formation of a stable enzyme-metal-CO<sub>2</sub>[14C]carboxyarabinitol biphosphate quaternary complex were unsuccessful indicating that correct assembly had not occurred (Gatenby, 1984; Somerville *et al.*, 1983; 1986). The formation of aggregated large subunit complexes may occur either during protein synthesis or can be a result of post translational events.

In contrast are the observations made for the wheat Rubisco large subunit synthesized in *E.coli*. At a high level of synthesis about 60% of the wheat large subunits is soluble (Bradley *et al.*, 1986). Analysis of these large subunits on sucrose density gradients showed that they migrate at a position indicating a Mr of approximately 750 000; however no enzymic activity was detected (unpublished observations). It may be that the insolubility of the maize large subunits and the failure of the wheat large subunits to assemble into a stable oligomeric complex indicates the requirement for small subunits, normally present in the chloroplast where assembly occurs.

The small subunits are synthesized as precursors in the cytosol of the plant cell and cross the chloroplast envelope prior to assembly (see Section 1.2.3 and Section 1.2.4). The observation that the assembled Rubisco holoenzyme contains only mature small subunits, and that isolated spinach mature small subunits assemble with cyanobacterial large subunit octamers *in vitro* with restoration of enzymic activity (Andrews & Lorimer, 1985), led researchers to assume that the small subunit presequence is probably only required for targeting to and transport across the chloroplast envelope membranes. For expression in *E.coli*, the cDNA of the wheat Rubisco small subunit precursor was modified in such a way that translation was initiated from the *lacZ* ribosome binding site to give a fusion protein with

either eleven (Bradley *et al.*, 1986; van der Vies *et al.*, 1986) or twelve (Gatenby *et al.*, 1987)  $\beta$ -galactosidase amino acids at the aminotermminus. In both these fusion proteins the first four amino acids of the mature small subunit had also been deleted.

When the wheat small subunit fusion protein is synthesized in *E.coli* in the absence of the large subunit the majority (80%) was detected in the soluble fraction, but was found to be unstable with a half-life of approximately 15 min. However when maize large subunits are synthesized in *E.coli* cells that contain a pool of wheat small subunits a larger portion of both the polypeptides is detected in the insoluble fraction (Gatenby *et al.*, 1987). Pulse-chase experiments showed that both the polypeptides are stable for more than 120 min and. This stability is perhaps due to the aggregation of the subunits. The failure to detect Rubisco activity indicates that the two subunits are not assembled correctly (Gatenby *et al.*, 1987). It should be noted that in these experiments the maize large subunit gene is under the control of the temperature sensitive bacteriophage lambda P<sub>L</sub> promoter and the possibility that the observed stability and possible incorrect association of the two subunits is the result of either the enhanced or reduced synthesis of certain *E.coli* proteins due to the increased temperature cannot be excluded.

It could be argued that the synthesis of homologous rather than heterologous Rubisco subunits in *E.coli* would create a more favourable situation in which assembly of a functional enzyme may occur; this possibility was examined by Bradley *et al.*, 1986). In these experiments both the wheat Rubisco large and small subunit genes were under the control of the *lac* promoter. The modified small subunit sequence was present on a multicopy plasmid, whereas the large subunit gene was present in bacteriophage M13, thus allowing differential expression of the two sequences. It was shown that when large subunits are synthesized in a pool of pre-existing small subunits both the subunits remain soluble. Analysis of cell extracts on sucrose density gradients revealed that, although the large subunits are distributed throughout the gradient, an abundant peak of large subunits sediments at a position indicating an apparent Mr of approximately 750 000. The small subunits are not associated with the large subunits and migrate near the top of the gradient, which suggests either a monomeric or dimeric form; no enzyme activity could be detected.

Thus even with a homologous set of subunits it is not possible to assemble a functional plant Rubisco enzyme in *E.coli*.

#### 1.4.2 SYNTHESIS AND ASSEMBLY OF PROKARYOTIC RUBISCO SUBUNITS

The prokaryotic nature of the bacterial Rubisco genes is likely to ensure correct transcription and translation in *E.coli*. Different forms of Rubisco have been synthesized and assembled in *E.coli*, among which is the most simple form of Rubisco i.e. the homodimeric large subunit enzyme from *Rhodospirillum rubrum* (Somerville & Somerville, 1984; Larimer *et al.*, 1986). This homodimeric Rubisco is encoded by a single gene which when expressed in *E.coli* produces large subunit polypeptides up to 12% of the total *E.coli* proteins. These polypeptides have the same apparent Mr as the purified *R.rubrum* dimer polypeptides, and assemble into enzymically active Rubisco (Somerville & Somerville, 1984). In these initial experiments the recombinant large subunits contained 25 aminoterminal amino acids from  $\beta$ -galactosidase. The kinetic properties of the recombinant dimeric enzyme were shown to be indistinguishable from the authentic *R.rubrum* enzyme, suggesting that the aminoterminal part of the polypeptide chain is positioned away from the active site of the enzyme. A high level of synthesis of the *R.rubrum* authentic large subunits, i.e. 19% of the total *E.coli* protein, is achieved when transcription of the gene is initiated using a *trp-lac (tac)* hybrid promoter plus the authentic ribosome binding site (Larimer *et al.*, 1986). The bacterium *Rhodopseudomonas sphaeroides* contains a Rubisco enzyme that probably consists of six large subunits (Gibson & Tabita, 1977), which are encoded by a single gene. When the large subunits are synthesized in *E.coli* they assemble into a functional enzyme molecule (Quivey & Tabita, 1984; Tabita *et al.*, 1986).

The ability to assemble an active Rubisco enzyme in *E.coli* allows researchers to study the properties of altered forms of Rubisco assembled from polypeptide chains containing specific mutations in known positions. A number of mutant Rubisco dimeric enzymes have been reported and are reviewed by Gutteridge *et al.* (1986) and Gutteridge & Gatenby, (1987). The Rubisco enzyme from the above mentioned purple non-sulphur

bacteria however are atypical in the sense that they do not contain small subunit polypeptides as does the enzyme found in plants.

Among the prokaryotes that contain a plant-like Rubisco are the cyanobacteria. The genes for the Rubisco large and small subunits in the cyanobacterium *Anacystis nidulans* (also called *Synechococcus*) are present in a single operon (Shinozaki & Sugiura, 1983; 1985). Gatenby *et al.*, (1985) were the first to study the synthesis and assembly of the Rubisco subunits from *Synechococcus* in *E.coli*. Since the cyanobacteria do not contain organelles, the small subunits do not contain a presequence and are synthesized as mature small subunit polypeptides. The expression of the *Synechococcus* Rubisco operon in *E.coli* was initially achieved by initiation of transcription from a bacteriophage lambda P<sub>L</sub> promoter, followed by temperature induced translation (Gatenby *et al.*, 1985). The detection of both the Rubisco subunits and Rubisco activity in *E.coli* extracts led to the conclusion that the cyanobacterial Rubisco subunits had been assembled into a functional protein complex. The apparent Mr of this enzymically active protein complex is approximately 550 000 and this suggests that possibly a fully assembled hexadecameric enzyme has been formed. However calculations of the composition of the assembled protein complex suggested a reduced amount of small subunits (probably 2 to 3 per large subunit octamer), which could explain the observed low specific activity. More recently it has been shown that the expression of the Rubisco operon using the *lac* promoter increases the amount of Rubisco synthesized and enables the assembly of hexadecameric structures (Bradley *et al.*, 1986; van der Vies *et al.*, 1986). The synthesis and assembly of hexadecameric cyanobacterial Rubisco in *E.coli* has also been reported by others (Christeller *et al.*, 1985; Gurevitz *et al.*, 1985; Tabita & Small, 1985), as well as for Rubisco from the purple sulphur bacterium *Chromatium vinosum* (Viale *et al.*, 1985) and the purple non-sulphur bacterium *Rhodopseudomonas sphaeroides* (Gibson & Tabita, 1986).

The eukaryotic chromophytic alga *Olithodiscus luteus* contains a plant-like Rubisco, but both the Rubisco genes are co-transcribed from a single operon present in the chloroplast genome (Reith & Cattolico, 1986). The prokaryotic nature of the *O.luteus* operon allowed a high level of Rubisco subunit synthesis when transcription was initiated using the *lac*

promoter. Under these conditions a small proportion of the subunits assembled into an active oligomer (Newman & Cattolico, 1988). The assembly of a functional hexadecameric Rubisco enzyme in *E.coli* provides the possibility to examine the mechanism of Rubisco assembly which may provide information to the function of the small subunit.

The deletion of the small subunit gene from the *Synechococcus* Rubisco operon allows the synthesis of the large subunits only. When large subunits are synthesized in the absence of small subunits no Rubisco activity could be detected in *E.coli* extracts (Christeller *et al.*, 1985; van der Vies *et al.*, 1986), suggesting that the small subunits may be essential for enzymic activity (see Section 1.2.2 and Section 1.2.4). This idea is supported by the observation that enzymic activity is restored when *E.coli* cells containing a small subunit deletion plasmid were infected with a bacteriophage M13 containing the small subunit gene (van der Vies *et al.*, 1986). The major soluble protein structures formed in *E.coli* cells synthesizing the large subunits only are large subunit dimers and octamers (van der Vies *et al.*, 1986). The large subunit dimers may represent an intermediate form in the pathway of the assembly of octameric large subunit structures. When both large and small subunits are synthesized in the same *E.coli* cell, the large subunit dimer is not observed, indicating that, although the small subunits are not essential for the formation of the large subunit octamer, they may reduce the pool of large subunit dimers and thus promote the formation of the octamer (van der Vies *et al.*, 1986).

It has recently been reported that partially purified Rubisco large subunit octamers possess approximately 1% of the catalytic activity of the holoenzyme (Andrews, 1988). This observation clearly shows that the small subunits are not absolutely essential for enzymic activity but act to enhance this activity, by a large factor, by possibly stabilizing the large subunit octameric structure (see also Fig. 2A).

#### 1.4.3 SYNTHESIS AND ASSEMBLY OF HETEROLOGOUS RUBISCO SUBUNITS

The assembly of a functional heterologous Rubisco enzyme was achieved when *E.coli* cells containing the cyanobacterial plasmid-encoded large subunit gene were infected

with a bacteriophage M13 that contained the wheat small subunit gene (van der Vies *et al.*, 1986). The specific activity of the hybrid Rubisco was about 10% of that of the wild-type, suggesting that either the cyanobacterial large subunits complex is not saturated with wheat small subunits or the failure of the wheat small subunits to enhance the activity to the level observed with the homologous subunits, possibly due to species diversity. Thus the observation that the wheat small subunits can assemble in *E.coli* with cyanobacterial large subunit octameric structures, but not with plant large subunits, suggests that the main requirement for the assembly of plant Rubisco may be the formation of a stable large subunit octamer which is suitable for association with the small subunits.

### 1.5 AIM OF THE PRESENT WORK

The initial aim of this project was to continue the study of the suggested role of the Rubisco large subunit binding protein in the correct assembly of the Rubisco subunits in chloroplasts, by isolating and characterizing cDNA for the binding protein subunits. It was planned to coexpress these sequences in the same *E.coli* cell that is expressing the Rubisco large and small cDNA sequences to determine whether this will rescue the assembly of plant Rubisco and so permit mutagenesis experiments to proceed.

## 2 MATERIALS AND METHODS



## 2.1 MATERIALS

### 2.1.1 BIOLOGICAL MATERIALS

#### 2.1.1.1 Plants

Seeds of *Pisum sativum* (var. Feltham First) were obtained from Charles Sharp, Boston Road, Sleaford, Lincolnshire and *Triticum aestivum* (var. Mercia) was a generous gift of the Institute for Plant Science Research, Cambridge.

#### 2.1.1.2 Animals

New Zealand white rabbits (female) were obtained from Rosemead Rabbits, Waltham Abbey, Essex.

### 2.1.2 NON-BIOLOGICAL MATERIALS

All of the materials used were of the highest analytical grade available. The source of specific chemicals and enzymes is listed below.

#### 2.1.2.1 Chemicals and reagents

Amersham International plc, Amersham, Buckinghamshire:  
 $\text{Na}_2\text{H}^{14}\text{CO}_3$  (specific activity 1.85 GBq/mmol = 50 mCi/mmol), [ $^{32}\text{P}$ ] CTP (specific activity  $\pm 110$  TBq/mmol =  $\pm 3,000$  Ci/mmol), [ $\alpha$ - $^{32}\text{P}$ ]dGTP (specific activity  $\pm 110$  TBq/mmol =  $\pm 3,000$  Ci/mmol), Na [ $^{125}\text{I}$ ] (specific activity 574 MBq/ $\mu\text{g}$  iodine = 15.5 mCi/ $\mu\text{g}$  iodine), [ $\alpha$ - $^{35}\text{S}$ ]dATP (specific activity  $\pm 44$  TBq/mmol = 1,200 Ci/mmol);

L-[<sup>35</sup>S]methionine (specific activity > 37 TBq/mmol = > 1,000 Ci/mmol), Amplify, bacteriophage  $\lambda$  DNA and the biotin-streptavidin system.

BDH chemicals Ltd., Atherstone, Warwickshire:

Amberlite monobed resin MB-3, APS, Hepes, 2-mercaptoethanol, ONGP, PEG6000, TEMED, Tris and 1,1,1-trichloroethane.

Bio-rad laboratories Ltd., Business park, Watford, Hertfordshire:

Protein assay kit.

Difco laboratories, Basingstoke, Hampshire:

Bacto-agar, bacto-tryptone and casamino acids.

Eastman Kodak, New York, U.S.A.:

N,N'-methylene bisacrylamide, N,N,N,N'-tetramethylene diamine.

Fisons plc, Loughborough, Leicestershire:

Acrylamide, formamide, glycerol and SDS.

Gibco Ltd., Paisley, Renfrewshire, Scotland:

Freund's incomplete adjuvant.

Oxoid Ltd., Basingstoke, Hampshire:

Bacto-yeast extract.

Pharmacia (United Kingdom) Ltd., Milton Keynes, Buckinghamshire:

ddATP, ddCTP, ddGTP, ddTTP,  $\gamma$ -methacryloxypropyltrimethoxysilane, percoll and m<sup>7</sup>G(5')ppp(5')G.

Pierce Chemical Company, Rockford, Illinois, U.S.A.:

Iodobeads.

Promega Biotec, P & S Biochemical Ltd, Liverpool, Merseyside:

Riboprobe Gemini System i.e. pGEM-blue vectors.

Sigma Chemical Co. Ltd, Poole, Dorset:

Agarose medium EEO (type II), ampicillin (sodium salt), ATP, BSA, chloroamphenicol 4-chloro-1-naphthol, Coomassie brilliant blue R250, creatine phosphate, cysteine-HCl, CTP, dATP, dCTP, dGTP, dTTP, DTT, DNA (salmon sperm), EtBr, GTP, IPTG, molecular weight kits Mw-SDS-blue and Mw-ND-500, Mops, ovalbumin, PMSF, PVP, RuBP (98% purity), spermidine 3 HCl and Tween 20.

Schleicher & Schull, Dassel, W-Germany:

Nitrocellulose type BA 85/1.

#### 2.1.2.2 Enzymes

Amersham International plc, Amersham, Buckinghamshire:

Nuclease S1, DNA polymerase I (Klenow fragment), DNA polymerase I (Kornberg), SP<sub>6</sub> RNA polymerase, T<sub>7</sub> RNA polymerase and restriction enzyme endonucleases.

BCL Boehringer Mannheim, Lewes, Sussex:

Aldolase, calf intestinal alkaline phosphatase (CIP) and exonuclease III.

BRL, Paisley, Renfrewshire, Scotland:

DNA T<sub>4</sub> ligase.

Progenia Biotec, P & S Biochemical Ltd, Liverpool, Merseyside:

RNasin.

Sigma Chemicals Co. Ltd, Poole, Dorset:

Aldolase, DNase I,  $\beta$ -galactosidase and RNase A.

## 2.2 METHODS

### 2.2.1 GROWTH AND MAINTENANCE OF BIOLOGICAL MATERIALS

#### 2.2.1.1 Growth of plants

Seeds of *Triticum aestivum* (var. Mercia) and *Pisum sativum* (var. Feltham First) were sown in J. Arthur Bowers potting compost and grown at  $20 \pm 2^\circ\text{C}$  for 7 days and 9 days respectively with a photoperiod of 12 h using "warm white" fluorescent lights (Philips), giving a light intensity of photosynthetically active radiation of  $50 \mu\text{E}, \text{m}^{-2}, \text{s}^{-1}$ . The seedlings were watered when required and the age of the seedlings taken from the time of sowing.

#### 2.2.1.2 Growth and storage of bacterial strains

##### 2.2.1.2.1 Growth of bacterial strains

*Escherichia coli* (*E.coli*) strain TG2 was grown in either Luria-Bertani (LB) medium or minimal glucose (MG) medium, as described in Section 2.2.2. When strain TG2 had been transformed with a plasmid the appropriate antibiotics were added to the medium, to maintain selection of the plasmid. For infection with bacteriophage M13, *E.coli* strain TG2 was grown in yeast-tryptone medium (YT), as described in Section 2.2.2. *E.coli* strain Y1090 was grown in LB medium containing 0.2% (w/v) maltose and ampicillin ( $50 \mu\text{g/ml}$ ).

##### 2.2.1.2.2 Storage of bacterial strains

*E.coli* bacterial strains were stored initially according to Maniatis *et al.* (1982). Short term storage (4-6 weeks) was either on LB or MG agar plates. The agar plates contained 1.5% (w/v) Difco-agar (Section 2.2.2.1). Plates were tightly wrapped in Parafilm and stored

inverted at 4°C. Long term storage was for longer than 2 months and up to 2-3 years. A 25 ml Universal vial containing 10 ml of liquid medium plus the appropriate antibiotics was inoculated with a single bacterial colony and incubated overnight at 37°C. Cells were harvested by centrifugation in a MSE Coolspin centrifuge at 3800 g for 10 min at 4°C and resuspended in 3-5 ml L-glycerol (2 g Difco-tryptone, 1 g Difco-yeast extract, 1 g NaCl, 100 ml glycerol, distilled water to 200 ml pH 7.2). Stocks were stored at -20°C.

#### 2.2.1.3 Growth and storage of bacteriophages

##### 2.2.1.3.1 Growth of bacteriophages

Bacteriophage  $\lambda$  was grown on *E.coli* Y1090 in either YT medium or CY medium (Section 2.2.2). Bacteriophage M13 was grown on a F' or Hfr strain of *E.coli* such as TG<sub>2</sub>, in YT medium (Section 2.2.2 and Section 2.2.4.2).

##### 2.2.1.3.2 Storage of bacteriophages

Bacteriophage  $\lambda$  was stored at 4°C in phage buffer (SM) which consisted of 50 mM Tris-HCl, 100 mM NaCl, 10 mM MgSO<sub>4</sub>, 0.01% (w/v) gelatine, pH 7.5, containing 0.5% (v/v) chloroform. Bacteriophage M13 was either stored on YT plates (Section 2.2.2.1) or as a liquid phage culture in YT medium at 4°C (Section 2.2.4.3).

#### 2.2.2 MEDIA AND ANTIBIOTICS

All media were autoclaved at 15 psi for 20 min.

Luria-Bertaini (LB) medium:

Bacto-tryptone 10g

Bacto-yeast extract 5g

NaCl 10g

distilled water to 1 litre

Yeast-tryptone (YT) medium:

Bacto-tryptone 10g

Bacto yeast extract 10g

NaCl 5g

distilled water to 1 litre

Minimal Glucose (MG) medium:

$\text{Na}_2\text{HPO}_4$  0.6 g

$\text{KH}_2\text{PO}_4$  0.3 g

NaCl 0.05 g

$\text{NH}_4\text{Cl}$  0.1 g

distilled water to 90 ml

adjusted to pH 7.4 and autoclaved. The following sterile solutions were then added:

20% glucose 1.0 ml

0.1 M  $\text{MgSO}_4$  0.1 ml

1 M  $\text{CaCl}_2$  0.1 ml

Casamino acid yeast (CY) medium:

casamino acids 10 g

bacto yeast extract 5 g

NaCl 3 g

KCl 2 g

distilled water to 1 litre

#### 2.2.2.1 Media containing agar or agarose

Liquid media was made up as described above (Section 2.2.2). Just before autoclaving one of the following was added:

For plates:	Difco-agar	15 g/l
	agarose	7 g/l

For top agar:	Difco agar	15 g/l
	agarose	7 g/l

#### 2.2.2.2 Antibiotic preparation

Ampicillin was dissolved in water at a concentration of 100 mg/ml, sterilized by membrane filtration and aliquots stored at -20°C. Ampicillin was added to the media to a final concentration of 100 µg/ml unless stated differently.

Chloroamphenicol was dissolved in 100% ethanol at a concentration of 34 mg/ml and stored in aliquots at -20°C.

#### 2.2.3 BACTERIAL TRANSFORMATION

Transformation of *E.coli* cells with plasmid DNA was performed using a modification of the method used by Mandel & Higa (1970). LB medium (10 ml) was inoculated with a single colony of the appropriate strain from a fresh plate and incubated overnight at 37°C. An aliquot of the overnight culture (500 µl) was used to inoculate 50 ml of LB medium and incubated further at 37°C until the absorbance (A) at 600 nm was 0.3 - 0.4. A sample of the cell culture (25 ml) was transferred to a Universal tube and centrifuged in a MSE 18 centrifuge at 3800 g for 10 min at 4°C. The cell pellet was resuspended in 10 ml ice-cold 0.1



M  $\text{MgCl}_2$  and respun as before. The cells were resuspended in 2 ml ice-cold 0.1 M  $\text{CaCl}_2$  and left at 4°C at least 4 h prior to transformation. After 12-24 h, the efficiency of transformation will increase fourfold to sixfold (Dagert & Ehrlich, 1979). A solution (0.1 ml) consisting of 3 volumes of SSC (0.15 M NaCl, 0.015 M Na citrate, pH 7.0) and 4 volumes of 0.1 M  $\text{CaCl}_2$ , was added to the DNA solution (volume not larger than 20  $\mu\text{l}$ ), and after mixing, 0.2 ml of competent *E.coli* cells were added and the mixture kept on ice for 15-30 min. The cell suspension was incubated in a waterbath of 42°C for 2 min and kept at 0°C for an additional 15-30 min. LB medium (0.7 ml) was added and the cells were allowed to grow at 37°C for 30 min without shaking. Samples of the transformation mixture (10-100  $\mu\text{l}$ ) were spread onto selective LB plates and grown at 37°C for 16-24 h. Individual colonies were used to prepare stocks (Section 2.2.1.2.2) and for the isolation of DNA (Section 2.2.5).

#### 2.2.4 BACTERIAL TRANSFECTION

##### 2.2.4.1 Bacteriophage $\lambda$

Phage  $\lambda$  was grown on *E.coli* strain Y1090. Plating cells were prepared as described in Maniatis *et al.* (1982) and were always used within 2 days of preparation. To plate the phage 0.1 ml of plating bacteria was incubated with an appropriate amount of  $\lambda$  phage to give:-

- 1) individual plaques i.e.  $10 - 10^2$  pfu (plaque forming units) per 85 mm Petridish or
- 2) confluent lysis i.e.  $10^4 - 10^5$  pfu per 8.5 mm Petridish.

The mixtures were incubated for 30 min at 37°C to allow the phage particles to adsorb to the bacterial cells. The suspension was added to 3 ml of LB top agar containing 0.2% (w/v) maltose, 10 mM  $\text{MgSO}_4$ , 50  $\mu\text{g/ml}$  ampicillin at 48°C. The solutions were mixed quickly and gently, and poured onto a LB agar plate containing 0.2% (w/v) maltose, 10 mM  $\text{MgSO}_4$ , 50  $\mu\text{g/ml}$  ampicillin. The plate was gently stirred to allow the distribution of the bacteria and the top agar. The plates were left at room temperature in order to allow the agar to set and then incubated at 37°C. After 12-16 hrs plaques could be seen. From plates

containing individual plaque the titre (pfu/ml phage stock) of the original phage stock was calculated. To prepare a phage stock individual plaques were picked as described in Maniatis *et al.* (1982) and the bacteriophage suspension stored as described in Section 2.2.1.3.2. Plates were carefully wrapped with Parafilm and stored at 4°C up to 10 days.

#### 2.2.4.2 Bacteriophage M13

Phage M13 was grown on *E.coli* strain TG<sub>2</sub>. Cells of strain TG<sub>2</sub> were streaked onto MG agar plates and after growth at 37°C an individual colony was picked to inoculate 10 ml YT medium (Section 2.2.2) which was subsequently incubated overnight at 37°C. An aliquot of the overnight culture (500 µl) was used to inoculate 5 ml YT medium and incubated at 37°C until the absorbance (A) at 600 nm was 0.8 - 1.0. The cells were used immediately; 0.2 ml of the cell culture and 10 µl of M13 phage stock of an appropriate titre i.e. 10-10<sup>2</sup> pfu per 85 mm Petridish; were added to 3 ml of molted top agar plus the required antibiotic at 48°C, mixed gently and poured onto a YT agar plate containing the antibiotic. The plate was left at room temperature in order for the agar to solidify and then incubated at 37°C for 16-24 h. Although the cells are not killed by the M13 phage, their growth is inhibited and the virus therefore produces turbid plaques on a *E.coli* cell lawn. From the number of plaques on the plate the exact titre (pfu/ml) of the phage stock was calculated. The plates were wrapped with Parafilm and stored at 4°C or individual plaques used to prepare a phage stock as described in Section 2.2.1.3.2.

#### 2.2.4.3 High-titre lysate preparation

Bacteriophage M13 was grown on a lawn of *E.coli* TG<sub>2</sub> cells to give individual colonies as described above. YT medium (10 ml) was inoculated with a single colony of TG<sub>2</sub> from a fresh MG agar plate and incubated overnight with shaking at 37°C or left on the bench at room temperature. The overnight TG<sub>2</sub> cell culture was diluted 100x in YT medium (10 ml) and a single M13-infected colony picked from a fresh plate and added to the YT medium.

Cells and phage were allowed to grow together for 3-4 h at 37°C in a shaker. The cells were harvested by centrifugation in a MSE, Coolspin centrifuge at 3800 g for 15 min at 4°C and the bacteria in the supernatant killed by incubation for 10 min at 70°C. The titre of the phagestock was determined as described above and the phage suspension stored at 4°C. A typical preparation gave a titre of  $10^{10}$  -  $10^{11}$  pfu/ml.

## 2.2.5 NUCLEIC ACID ISOLATION

### 2.2.5.1 DNA extraction and precipitation

#### 2.2.5.1.1 DNA extraction with phenol/chloroform

Phenol was equilibrated with 0.1 M Tris-HCl pH 8.0 as described in Maniatis *et al.* (1982). All the "phenol" mentioned in this thesis will refer to equilibrated phenol. The "chloroform" is a mixture of chloroform and isoamyl alcohol (24:1, v/v) and will be referred to as chloroform. To extract the DNA from a given solution an equal volume of phenol/chloroform mixture (1:1) was added and mixed by vortexing. The mixture was centrifuged in an Eppendorf, Micro Centaur (microfuge) at 11 600 g for 5 min at room temperature. A pipette was used to transfer the aqueous phage (top-layer) to a fresh Eppendorf tube avoiding the interface and lower organic phase.

#### 2.2.5.1.2 DNA precipitation with ethanol

DNA precipitation was performed according to Maniatis *et al.* (1982). Samples were stored either at -20°C for 30-60 min or at -70°C for 15 min; Na acetate and ethanol were used to precipitate the DNA unless stated differently. The DNA pellet was washed with 70% ethanol at -20°C prior to drying under vacuum and dissolved in either TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0) or T E<sub>0.1</sub> buffer (10 mM Tris-HCl, 0.1 mM EDTA pH 8.0). The DNA was stored at 4°C (1-2 weeks) or at -20°C (longer than 2 weeks).

### 2.2.5.2 Large-scale isolation of plasmid DNA

#### 2.2.5.2.1 Large-scale isolation of plasmid DNA on CsCl gradients

Plasmid DNA was isolated from *E.coli* by a modification of the alkaline lysis method of Birnboim and Doly (1979). Ten ml of LB medium containing the appropriate antibiotic in a 25 ml Universal container was inoculated with a single bacterial colony from a fresh LB plate. The culture was grown overnight at 37°C in a orbital shaker at 200 rpm. A two litre flask containing 800 ml of LB medium plus antibiotic was inoculated with 8 ml of the overnight culture and incubated at 37°C with shaking until the cell suspension had an absorbance (A) at 600 nm of 1.0. The DNA was amplified by adding chloramphenicol, a inhibitor of protein synthesis, to a final concentration of 100 µg/ml. The cell culture was incubated for a further 16 h at 37°C in an orbital shaker. Cells were harvested by centrifugation in a MSE Mistral 6L centrifuge at 2075 g for 20 min at 4°C. Cells were resuspended in 5 ml ice-cold TS buffer (50 mM Tris-HCl, 25% (w/v) sucrose, pH 8.0) and the suspension transferred to 50 ml Oakridge tubes which were kept in ice-water. One ml of TS buffer containing lysozyme in a concentration of 5 mg/ml was added and the mixture left in ice-water for 10 min. EDTA (2.8 ml at 250 mM, pH 8.0) was added to the tube which was incubated at 0°C for another 5 min with occasional swirling by hand. After addition of 8.8 ml TET buffer (50 mM Tris-HCl, 60 mM EDTA, 10% (v/v) Triton X-100, pH 8.0) and gentle mixing, the solution was centrifuged in a MSE18 centrifuge at 1100 g for 60 min at 4°C.

The supernatant containing the DNA was transferred to a clean Oakridge tube and 6 ml of 30% (w/v) polyethylene glycol-6000 (PEG-6000), 1.5 M NaCl was added; the content of the tube was gently mixed and the tube was left at 0°C for 15 min. The DNA was collected by centrifugation in a MSE18 centrifuge at 48 000 g for 15 min at 4°C. The pellet was drained and 2 ml of TE buffer was added to dissolve the DNA completely.

The plasmid DNA was further purified on CsCl density gradients (Maniatis *et al.*, 1982). CsCl (3.218 g) was dissolved in the 2 ml DNA solution and 0.4 ml ethidium bromide (EtBr) at 10 mg/ml, added. The final volume was adjusted to 4.2 ml with TE buffer. In order to separate contaminating proteins, the DNA solution was centrifuged in a MSE 18 centrifuge at 12 000 g for 15 min at room temperature. The proteins pelleted at the bottom and on the sides of the centrifuge tube. The DNA solution was carefully transferred to a small 5 ml centrifuge tube. The solution was overlaid with 0.2 ml of liquid paraffin. The gradients were formed by centrifugation in a Beckman, L8-70 M ultracentrifuge in a vertical rotor type 50 Ti spun at 50,000 rpm for 16 h at 20°C. The rotor was allowed to come to rest without the aid of the brake. The plasmid band was visualized under ultraviolet (u.v.) light and collected from the tube by the use of a 18 gauge hypodermic needle inserted into the side of the tube. The collected volume was kept under 2 ml. The EtBr was removed from the DNA solution by extraction with 4 ml of isopropanol saturated with SSC (0.15 M NaCl, 0.015 M Na-acetate). The extraction was repeated 4-5 times. Removal of isopropanol and CsCl was accomplished by extensive dialysis against TE buffer, 3 changes of 1 litre each. The last traces of proteins were removed from the DNA preparations by phenol extraction (Section 2.2.5.1.1) and the plasmid DNA was precipitated with ethanol (Section 2.2.5.1.2). The plasmid DNA was stored at 4°C (1-2 months) or -20°C (> 2 months). The yield of an average preparation was 0.5 - 1 mg of DNA.

#### 2.2.5.2.2 Rapid large-scale isolation of plasmid DNA

The method used was the one described in the Technical Bulletin of Promega Biotec (1986). This method is much quicker and gives a higher yield than the one described in Section 2.2.5.2.1. This rapid large-scale method became available during the course of this work.

LB medium (10 ml) containing the appropriate antibiotic was inoculated with a single bacterial colony from a fresh LB plate, and incubated overnight at 37°C in an orbital shaker. A one litre flask containing 250 ml of LB medium plus antibiotic was inoculated with 2.5 ml

of the overnight culture and incubated at 37°C overnight in an orbital shaker at 200 rpm. The cells were harvested by centrifugation in a MSE18 centrifuge at 4000 g for 10 min at 4°C and thoroughly resuspended in 6 ml of freshly prepared isolation medium consisting of 25 mM Tris-HCl, 10 mM EDTA, 15% (w/v) sucrose and 2 mg/ml lysozyme, pH 8.0 and left in ice-water for 20 min. NaOH (12 ml of 0.2 M) containing 1% (w/v) sodium dodecyl sulphate (SDS) was added and the mixture carefully and thoroughly mixed by inversion and incubated in icewater for 10 min. Na acetate (7.5 ml of 3M at pH 4.6) was added, the solution mixed by gentle inversion and incubated at 0°C for 20 min. The DNA was separated from the cell debris by centrifugation in a MSE 18 centrifuge at 27 000 g for 15 min at 4°C. The supernatant was transferred to a clean tube and incubated at 37°C for 20 min with 50 µl RNase (1 mg/ml). Phenol/chloroform extraction was performed twice, as described in Section 2.2.5.1.1.

The DNA was precipitated with 2 volumes of ethanol (see Section 2.2.5.1.2) collected by centrifugation in a MSE18 centrifuge at 5000 g for 15 min at 4°C and dissolved in 1.6 ml of distilled water. NaCl (0.4 ml of 4 M) was added and the solution mixed, after which 2 ml of 13% (w/v) PEG-6000 was added and the solution mixed again before incubation in ice-water for 1 h. The precipitated plasmid DNA was harvested by centrifugation in a MSE centrifuge at 12 000 g for 10 min at 4°C. The pellet was washed with 70% ethanol (-20°C) and dried under vacuum before being dissolved in 0.5 ml of TE buffer. This procedure yields 0.5-0.7 mg of DNA.

### **2.2.5.3 Small scale plasmid DNA isolation**

#### **2.2.5.3.1 Rapid isolation of plasmid DNA**

The method used was initially developed by Birnboim and Doly (1979) and is based on the alkaline lysis protocol. An aliquot of LB medium (2 ml) with the appropriate antibiotic was inoculated with a single colony from a fresh LB plate and grown overnight at 37°C. An aliquot of the overnight culture (1.5 ml) was transferred to an Eppendorf tube and

centrifuged in an Eppendorf microfuge at 11.600 g for 1 min at room temperature. The supernatant was discarded and the cell pellet resuspended in 0.2 ml lysis solution consisting of 25 mM Tris-HCl, 10 mM EDTA, 50 mM glucose, 2 mg/ml lysozyme (freshly dissolved), pH 8.0 and left in ice-water for 30 min. NaOH (0.4 ml of 0.2 M containing 1% (w/v) SDS was added and the mixture left at 0°C for 5 min with occasional mixing by inversion. After addition of 0.3 ml of 3 M Na acetate pH5, the mixture was incubated for 60 min at 0°C. The DNA was separated by centrifugation in an Eppendorf microfuge at 11 600 g for 5 min at 4°C. The supernatant (0.5 ml) was transferred to a fresh microfuge tube and the DNA precipitated with 1 ml ethanol (see Section 2.2.5.1.2). After centrifugation in an Eppendorf microfuge at 11 600 g for 5 min at 4°C, the supernatant was discarded and the DNA pellet dissolved in 0.2 ml of 0.1 M Na acetate, pH 6.0. Addition of 0.4 ml of ethanol was used to re-precipitate the DNA (see Section 2.2.5.1.2). The DNA pellet collected after centrifugation in an Eppendorf microfuge at 11 600 g for 5 min at 4°C, was washed with 0.5 ml of 70% (v/v) ethanol (-20°C), dried under vacuum and dissolved in 40 µl of TE buffer. This method yielded 1-1.5 µg of DNA.

#### 2.2.5.3.2 Isolation of highly purified plasmid DNA

Plasmid DNA was isolated from *E.coli* by a modification of the boiled lysis method described by Holmes and Quigley (1981). This method was used to isolate highly purified plasmid DNA in relatively large amounts which was initially used for DNA plasmid sequencing (Section 2.2.11). LB medium (10 ml) containing the appropriate antibiotic was inoculated with a single colony from a fresh LB plate and incubated at 37°C in an orbital shaker at 200 rpm overnight (i.e. 16 h). The cells were harvested by centrifugation in a MSE Coolspin at 3800 g for 10 min at room temperature. The cell pellet was drained and resuspended in 100 µl of TS buffer, transferred to a clean Eppendorf tube and kept at room temperature. M-STET buffer (600 µl of 50 mM Tris-HCl, 50 mM EDTA, 5% (w/v) sucrose, 5% (v/v) Triton X-100, pH 8.0 plus 7 µl lysozyme (50 mg/ml in 50 mM Tris HCl pH 8.0, freshly dissolved), was added to the cellsuspension which was gently mixed by inversion

and boiled immediately for 60 s. For more than two samples a M-STET /lysozyme mix was made and 600  $\mu$ l of this mix added to each cellsuspension. The boiled samples were centrifuged in an Eppendorf microfuge at 11 600 g for 45 min at room temperature until the pellet was thight.

The pellet was pulled out with a tooth pick and 1  $\mu$ l RNase A (10 mg/ml) was added to the remaining solution which was than incubated in a 37°C waterbath for 20 min. Two phenol/chloroform extractions were performed in the following way: 150  $\mu$ l equilibrated phenol (Maniatis *et al.*, 1982) was added and the solution was mixed on a Whirlimix, than 150  $\mu$ l chloroform (Section 2.2.5.1.1) was added to the solution and the DNA extracted by mixing on a vortex for 1 min (for details on centrifugation conditions see Section 2.2.5.1.1). After the second extraction the aqueous phase was transferred to a clean Eppendorf tube and one-tenth volume of 3 M Na acetate, pH 6, was added. The DNA was precipitated by adding an equal volume of isopropanol followed by incubation at -20°C for 30 min. The DNA was collected by centrifugation in an Eppendorf microfuge at 11 600 g for 10 min at room temperature and dissolved in 100  $\mu$ l TE buffer. Na acetate (10  $\mu$ l of 3 M, pH 6) was added and the DNA was re-precipitated with 2 volumes of ethanol as described in Section 2.2.5.1.2. The washed and dried DNA pellet was dissolved in 50  $\mu$ l TE<sub>0.1</sub> buffer. The yield was up to 25-30  $\mu$ g of plasmid DNA.

#### 2.2.5.4 Bacteriophage $\lambda$ DNA isolation

Phage DNA was isolated by a modification of the method described by Maniatis *et al* (1982). LB medium (10 ml) containing 0.2% (w/v) maltose and 50  $\mu$ g/ml ampicillin, was inoculated with a single colony of *E.coli* Y1090 from a fresh plate and incubated at 37°C overnight. An aliquot of the culture (0.25 ml) was diluted into 30 ml of  $\lambda$  phage buffer (3 g KH<sub>2</sub>PO<sub>4</sub>, 7 g Na<sub>2</sub>HPO<sub>4</sub>, 5 g NaCl, 1 ml 1M MgSO<sub>4</sub>, 1 ml 0.1 M CaCl<sub>2</sub>, 1 ml 1% (w/v) gelatin, distilled water to 1 litre), and 1 ml was inoculated with a single plaque from a fresh plate and incubated at 37°C for at least 20 min in a shaker. CY mixture (12 ml) consisting of 200 ml CY medium (Section 2.2.2) plus 6 ml Tris-HCl pH 7.4 and 0.6 ml 1 M MgSO<sub>4</sub> was



was added and the cell suspension incubated at 37°C for 7-10 h, until the cell suspension cleared. After the lysed cell culture was incubated for an additional 2 h, chloroform (0.5 ml) was added and the mixture stored at 4°C overnight. An aliquot of the lysate (1-2 ml) was saved in a sterile tube to which 2 drops of chloroform were added, and the tube stored at 4°C; this tube acted as a stock. The remainder of the lysate was transferred into 10 ml Corex tubes and centrifuged in a MSE 18 centrifuge at 12 000 g for 10 min at 4°C. The supernatant was transferred to a 30 ml Corex tube and 5 µl DNase I (10 mg/ml) and 5 µl RNase A (10 mg/ml) added. Incubation for at least 1 h at 37°C was followed by addition of 10 ml of 2.5 M NaCl, 20% (w/v) PEG (6000). The solution was mixed and kept at 4°C for at least 30 min. The phage DNA was collected by centrifugation in a MSE18 centrifuge at 6000 g for 10 min at 4°C. The supernatant was discarded and the pellet respun after which the remaining supernatant was drawn off and the pellet dissolved in 0.25 ml TE buffer. The DNA solution was quickly transferred to Eppendorf tubes and extracted with phenol (Section 2.2.5.1.1). The DNA was precipitated by addition of 0.5 ml ethanol (stored at -20°C); the solution was mixed and left at 0°C for 30 min, after which it was centrifuged in an Eppendorf microfuge at 11600 g for 2 min at 4°C. The DNA pellet was dissolved in 0.25 ml of TE buffer and ethanol precipitated again (Section 2.2.5.1.2). The final pellet was dried under vacuum and dissolved in 20 µl of TE buffer. Yields of 1-2 µg of DNA were obtained.

## 2.2.6 ISOLATION OF cDNA SEQUENCES

### 2.2.6.1 **Screening of a $\lambda$ gt11 expression library with antibodies**

The  $\lambda$  gt11 expression library was prepared by Dr. Christine Raines of the Institute for Plant Science Research, Cambridge. *Triticum aestivum*, variety Chinese Spring, was grown in a glass house under natural conditions in July 1986 at a temperature of 16-23°C. The wheat plants were grown on

compost and well watered. After 10 days the middle Section of the well developed leaves was used to isolated polyA<sup>+</sup> RNA which was subsequently used to prepare an expression library in bacteriophage  $\lambda$  gt11 as described by Young & Davies (1983 a, b). The amplified wheat  $\lambda$  gt11 expression library had a titre of  $10^9$ - $10^{11}$  pfu/ml.

The library was plated onto Y1090 *E.coli* cells on LB agarose plates containing ampicillin (50  $\mu$ g/ml) at a density of  $10^5$  pfu per 150 mm Petridish as described in Section 2.2.4.1. LB top agarose (9 ml), containing ampicillin (50  $\mu$ g/ml) and 0.3 ml of plating bacteria were used. When the top agarose was set, the plates were incubated at 42°C for approximately 3 $\frac{1}{2}$  h until plaques could be seen. Plates were overlayed with a labelled nitrocellulose filter that had been soaked in filter-sterile 10 mM IPTG (isopropylthiogalactoside) and dried on Whatman 3 MM filter paper. Plates were incubated for an additional 4 h at 37°C. The nitrocellulose filters were carefully taken off the plates which were wrapped in Parafilm and stored at 4°C. The filters were transferred to TBS buffer consisting of 50 mM TrisCl, 150 mM NaCl, pH 8.0 and washed with shaking, 3 times for 30 min each. After the filters were blocked in TBS buffer containing 3% (w/v) Marvel (dried milk) for 2 h, at 37°C in a shaker, they were rinsed twice with in TBS buffer and incubated with purified antibody (see Section 2.2.13.2.1) at 37°C overnight with shaking. The antibody solution was drained off and stored at -20°C in order to be re-used and the filters were washed for three 20 min periods with TBS on a shaker at room temperature. The bound antibody was detected as described in Section 2.2.14.2.2. Good autoradiographic signals were usually obtained after 3-5 days exposure.

#### 2.2.6.2 Purification of plaques

In the first round of screening, the  $\lambda$  gt11 library was plated at high density such that the plates would show confluent lysis of the *E.coli* lawn. By lining up the autoradiograph and the plate, an "area" of lysed cells could be picked as described by Maniatis *et al.* (1982) and a phage stock prepared as described in Section 2.2.1.3.2. A phage stock like this would yield  $10^8$ - $10^{10}$  pfu/ml. The phage was plated again but at a lower density ( $10^3$ - $10^4$  pfu per

150 mm Petri dish) and the screening procedure repeated as described before. The screening procedure was then repeated at an even lower density ( $10^2$ - $10^3$  pfu per 150 mm Petridish) and repeated as many times as needed until all the individual plaques on the plate gave a signal with the antibody. The purified phage was picked as described by Maniatis *et al.* (1982) and the phage suspension kept as a stock at 4°C.

## 2.2.7 NUCLEIC ACID RESTRICTION AND MODIFICATION

### 2.2.7.1 Restriction endonuclease digestion

DNA restriction enzyme digestion was performed according to the manufacturers instructions with the following modification: 0.5 - 1 µg of DNA in a volume of 1-2 µl TE buffer was digested with 3-5 enzyme units per µg of DNA in a final reaction mixture of 10-20 µl. The digestion was carried out in a 37°C waterbath for 60 min. If the digested DNA was to be analyzed by gel electrophoresis then gel loading buffer was added (see Section 2.2.12.1.1). The DNA was purified by phenol/chloroform extraction and precipitation with ethanol (see Section 2.2.5.1) if further enzymic modifications were required.

### 2.2.7.2 Dephosphorylation of DNA

The terminal 5'phosphate groups were removed from digested plasmid DNA by treatment with calf intestinal alkaline phosphatase (CIP) by means of a method modified from that originally described by Chaconas & van de Sande (1980). Digested DNA (1-5 µg) was dissolved in TE buffer and added to 180 µl of distilled water. An aliquot (20 µl) of 10x CIP buffer consisting of 0.5 M Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM ZnCl<sub>2</sub> and 10 mM spermidine, pH 9.0, was added and the sample mixed. The reaction was started by adding 1 µl of CIP (= 1 unit) and continued by incubation at 37°C for 15 min. The CIP was inactivated by heating at 60°C for 15 min, after which the sample was left to cool to room temperature. Fresh CIP (1 µl) was added and the sample incubated again at 37°C for 15

min. The reaction was stopped by adding 200  $\mu$ l 1 mM EDTA and the enzyme inactivated by heating at 60°C for 15 min. The DNA was extracted three times with 400  $\mu$ l phenol (see Section 2.2.5.1.1) and the final aqueous phase was extracted four times with an equal volume of ether. To remove the final traces of ether the open Eppendorf tube was placed in the fumehood to allow the ether to evaporate. An aliquot (10  $\mu$ l) of 10 mM LiCl was added and the DNA precipitated with two volumes of ethanol. After incubation at -70°C for 30 min the DNA was collected by centrifugation in an Eppendorf microfuge at 11 600 g for 10 min at 4°C. The DNA pellet was washed in 70% ethanol (stored at -20°C) and the vacuum-dried pellet dissolved in 20  $\mu$ l TE buffer. For analysis on an agarose gel (see Section 2.2.12.1.2) 2  $\mu$ l of the final preparation was used. Dephosphorylated DNA was stored at -20°C.

#### 2.2.7.3 DNA ligation

Linearized plasmid DNA which had been dephosphorylated as described in Section 2.2.7.2 and a desired DNA fragment were ligated using T<sub>4</sub> ligase, by means of a method modified from that described by Rusche & Howard-Flanders (1985). Linearized dephosphorylated plasmid DNA (100-200 ng) in TE buffer was mixed with 100-200 ng of an isolated fragment of DNA (see Section 2.2.8). An aliquot (1.5  $\mu$ l) of 10x ligation buffer, consisting of 0.5 M Tris-HCl, 0.1 M MgCl<sub>2</sub>, 0.01 M adenosine triphosphate (ATP), 0.01 M hexamine cobalt chloride (HCC) and 1 mg/ml bovine serum albumine (BSA), pH 7.5, was added to the DNA mixture. This was followed by the addition of an aliquot (1.5  $\mu$ l) of 50 mM dithiothreitol (DTT) and 14  $\mu$ l of distilled water. At this stage an aliquot (2  $\mu$ l) was often removed from the reaction mixture for later analysis by gel electrophoresis. The enzyme T<sub>4</sub> DNA ligase (1  $\mu$ l containing 2.5 units) was added, the sample mixed and incubated at 16°C for 16-20 h. After incubation a second aliquot (2  $\mu$ l) was often removed for analysis. Both aliquots were checked for successful ligation by electrophoresis through an agarose gel (see Section 2.2.12.1.2). An aliquot (2.5  $\mu$ l) of the remaining sample was used to transform bacteria to antibiotic resistance (see Section 2.2.3).

#### 2.2.7.4 Filling recessed 3' ends of double-stranded DNA

The method used was a modification of the one described by Maniatis *et al.* (1982). Plasmid DNA was digested with the desired restriction enzyme (Section 2.2.7.1) and analyzed on an agarose gel (Section 2.2.12.1.2) to confirm linearisation. To 10  $\mu$ l of the digestion mixture containing 1  $\mu$ g of DNA, 2  $\mu$ l 2 mM dNTPs (deoxyribonucleoside triphosphates, i.e. dATP, dCTP, dGTP, dTTP) prepared as described by Maniatis *et al.* (1982), 2  $\mu$ l of 10x NT (nick-translation) buffer, consisting of 0.5 M Tris-HCl, 0.1 M MgSO<sub>4</sub>, 1 mM DTT, 500  $\mu$ g/ml BSA, pH 7.2, 5  $\mu$ l distilled water and 1  $\mu$ l Klenow fragment of DNA polymerase (4 units per  $\mu$ l), were added and the reaction mixture incubated for 30 min at room temperature. The reaction was terminated by adding 2  $\mu$ l of 0.5 M EDTA pH 8.0, and the DNA extracted with 30  $\mu$ l of phenol/chloroform, followed by an additional extraction using 60  $\mu$ l chloroform (Section 2.2.5.1.1). Na acetate (5  $\mu$ l of 3M, pH6) was added and the DNA precipitated with 110  $\mu$ l ethanol stored at -20°C. (Section 2.2.5.1.2). The final DNA pellet was dissolved in 10  $\mu$ l TE buffer and 2  $\mu$ l used for religation (Section 2.2.7.3).

#### 2.2.7.5 Removal of 5' mononucleotides from double-stranded DNA

The enzyme exonuclease III from *E.coli* catalyzes the stepwise 3'  $\rightarrow$  5' removal of 5' mononucleotides from double-stranded DNA carrying a 3'-OH end (Weiss, 1976). The method used was essentially described by Henikoff (1984). Plasmid DNA (10  $\mu$ g) was digested to completion with an appropriate restriction enzyme (Section 2.2.7.1) and the DNA precipitated with ethanol (Section 2.2.5.1.2) and dissolved in TE<sub>0.1</sub> buffer to a final concentration of 200 ng/ $\mu$ l. To 50  $\mu$ l of the DNA solution, 6  $\mu$ l of 10x ExoIII buffer consisting of 0.5 M Tris-HCl, 50 mM MgCl<sub>2</sub>, 10 mM DTT, pH 8.0, and 4  $\mu$ l TE<sub>0.1</sub> buffer, were added. The mixture was incubated at 37°C for 2 min; the reaction was started by adding 1  $\mu$ l (100 units) of exonuclease III and a 5  $\mu$ l sample was removed as quickly as possible into 5  $\mu$ l of TE<sub>10</sub> (10 mM Tris-HCl, 10 mM EDTA, pH 8.0) which was kept in

ice-water. Samples were removed at desired intervals, i.e. 40 sec. When all the samples had been taken they were heated at 65°C for 10 min and kept in ice-water. The samples were always treated immediately with nuclease S1 (Section 2.2.7.6).

At the concentrations of DNA and exonuclease III described here, the enzyme would remove about 200-300 bases/min. The initial rate of digestion is very roughly proportional to the relative amounts of DNA and enzyme. The digestion rate could be controlled by performing the reaction at different temperatures. At 10 µg of DNA and 100 units of exonuclease III about 120-170 bases/min were removed at 30°C and 50-75 bases/min at 20°C (G. Murphy, personal communication).

#### 2.2.7.6 Degradation of single-stranded DNA

The enzyme nuclease S1 degrades single-stranded DNA (Vogt, 1973) to yield 5' phosphoryl mono- or oligonucleotides whereas double-stranded DNA is resistant to the enzyme. A series of pilot scale reactions to determine the amount of nuclease S1 required for a given amount of DNA is described by Maniatis *et al.* (1982). The method used was as follows. An aliquot (100 µl) of S1 nuclease buffer, consisting of 30 mM Na acetate, 250 mM NaCl, 1 mM ZnSO<sub>4</sub>, 100 µg/ml tRNA, pH 4.6 containing 150 U/ml nuclease S1, was added to the 10 µl DNA preparations which had been treated with exonuclease III (Section 2.2.7.5), and the samples kept at 0°C until the whole series was completed. After incubation at 37°C for 20 min, the reaction was terminated by adding 2 µl of 0.5 M EDTA pH 8.0, and the DNA precipitated with 200 µl ethanol (-20°C) and stored at -20°C for 40 min (Section 2.2.5.1.2). The final DNA pellet was dissolved in 10 µl TE buffer or in 10 µl DNA polymerase digestion mixture if it was to be treated with DNA polymerase, as described in Section 2.2.7.4.

#### 2.2.7.7 Radiolabelling of DNA

The method employed was modified from that described by Rigby *et al.* (1977) and involved the nick-translation of DNA with DNA polymerase I. The incubation was carried out using 0.25-0.5  $\mu\text{g}$  of isolated DNA fragment (Section 2.2.8) in TE buffer. An aliquot (5  $\mu\text{l}$ ) of 4 x NT buffer consisting of 200 mM Tris-HCl, 20 mM  $\text{MgCl}_2$ , 20  $\mu\text{g/ml}$  BSA, pH 7.5, was added followed by addition of 2  $\mu\text{l}$  dNTP mix which consisted of 200  $\mu\text{M}$  dATP, 200  $\mu\text{M}$  dCTP, 200  $\mu\text{M}$  TTP. Radioactively labelled dGTP was added (5  $\mu\text{l}$  of a 3.3 mM  $\alpha$ - $^{32}\text{P}$ dGTP solution with a specific activity of  $\pm 110 \text{ TBq/mmol} = \pm 3,000 \text{ Ci/mmol}$ ) followed by distilled water up to 19  $\mu\text{l}$ . The reaction mixture was placed in a  $15^\circ\text{C}$  waterbath and 1  $\mu\text{l}$  of 1 x DNase I mix added which was prepared as follows. DNA polymerase I was dissolved in distilled water at a concentration of 1 mg/ml and 4  $\mu\text{l}$  of Pol I buffer consisting of 2 ml BSA at 1 mg/ml, 100  $\mu\text{l}$  of 2 M  $(\text{NH}_4)_2\text{SO}_4$ , 20  $\mu\text{l}$  of 1 M 2-mercaptoethanol, 2 ml glycerol, was added. This solution served as a 100x stock. For 1x DNase I mix the stock solution was diluted 100x in Pol I buffer and stored at  $-20^\circ\text{C}$ ; Pol I buffer (0.25  $\mu\text{l}$ ) was added and the reaction mixture incubated for 2 h. The reaction was stopped by adding 10  $\mu\text{l}$  of stop buffer which consisted of 0.2 M Na acetate, 0.01 M Tris-HCl, 0.01 M EDTA, 0.5% (w/v) SDS, pH 7.5.

The protocol used to determine the percentage of [ $\alpha$ - $^{32}\text{P}$ ]dGTP incorporated into the DNA, was that described by Maniatis *et al.* (1982). An aliquot of the reaction mixture (2  $\mu\text{l}$ ) was spotted onto the centre of each of two 2.4 cm discs of Whatman DE81 filterpaper and one of the discs was washed in 0.5 M  $\text{Na}_2\text{HPO}_4$  for 6 periods of 5 min each, followed by once in distilled water for 1 min and then twice in 95% ethanol for 1 min each. All the washes were performed under constant shaking. The washed filter was allowed to dry, 4 ml of liquid scintillation fluid (LKB, 'Optiphase Safe') added and the radioactivity on the filter determined by means of a scintillation counter (LKB 1219 Rackbeta).

From the remainder of the reaction mixture, the labelled DNA was purified by separation from the unincorporated dNTPs using a Sephadex G50, superfine gelfiltration column (1 ml) in TE buffer. Sometimes a dye solution (2  $\mu\text{l}$ ) consisting of 0.1% (w/v) bromophenol blue,

10% (w/v) Ficoll (type 400) was added to the DNA solution before purification on the gelfiltration column in order to visualize the position of the free dNTPs on the column. The DNA was eluted from the column in TE buffer and 200  $\mu$ l fractions were collected. An aliquot of each fraction (1  $\mu$ l) was counted as described before. This procedure gave DNA with a high specific activity of about  $10^7$  cpm/ $\mu$ g of DNA which was used immediately for DNA hybridization (Section 2.2.10).

#### 2.2.8 ELECTRO-ELUTION OF DNA FROM AGAROSE GEL

DNA was digested (Section 2.2.7.1) such that the fragment of interest was released, and the digestion mixture analyzed on an agarose gel containing ethidium bromide (Section 2.2.12.1.2). After the gel was run, the mobility of the fragment was determined by examination of the gel under u.v. light. The gel was run such that the fragment to be isolated was well separated from other fragments to reduce contamination. The method used to isolate the DNA fragment from the gel was as described by Maniatis *et al.* (1982), with the following modifications. The DNA fragments (size range 1.1 - 2.1 kb) were electroeluted in 1x TBE buffer for 45 min at 60 mA. The process of electroelution was followed by examination of the gel at regular intervals under u.v. light. The isolated DNA fragment was purified by two phenol extractions and one phenol/chloroform extraction (Section 2.2.5.1.1), precipitated with ethanol (Section 2.2.5.1.2) and finally dissolved in TE buffer. The DNA was analyzed on an agarose gel before using it for subcloning (Section 2.2.7.3) or radioactive labelling (Section 2.2.7.7).

#### 2.2.9 BACTERIOPHAGE LAMBDA PLATE LYSATE PREPARATION

The procedure used to prepare a plate lysate was the one described as method 1 by Maniatis *et al.* (1982), with the modification that agarose was used instead of agar and no chloroform was added to the final supernatant.



#### 2.2.10 NUCLEIC ACID HYBRIDISATION

For the hybridisation of cDNA sequences isolated from the  $\lambda$  gt11 library a method based on that described by Southern (1975) was used. The isolated DNA was digested with the appropriate restriction enzyme (Section 2.2.7.1) and subjected to electrophoresis through an 1% (w/v) agarose gel in TAE buffer (Section 2.2.12.1.1). Phage  $\lambda$  DNA, digested with *Eco*R1 and *Hind*III (Section 2.2.7.1) was used to provide size range markers on the gel. After electrophoresis the DNA was visualized (Section 2.2.12.1.1) and a photograph taken, after which the gel was soaked in 0.25 M HCl for 10 min, with shaking. The DNA was denatured by incubating the gel in a buffer consisting of 1.5 M NaCl and 0.5 M NaOH, for two periods of 30 min each at room temperature with constant shaking. Neutralisation was achieved by soaking in neutralisation buffer consisting of 0.5 M Tris-HCl, 1.5 M NaCl, pH 7.0, for 30 min at room temperature. After two additional 30 min washes in neutralisation buffer the DNA was transferred to nitrocellulose. The method used was diffusion blotting in 10x SSC for 20-24 h as described by Maniatis *et al.* (1982). After transfer, the filter was rinsed in 2x SSC for 10 min at room temperature with shaking, and baked in a vacuum oven at 80°C for 2 h. At this stage the filter can be stored indefinitely.

The hybridization rates and DNA hybrid stability are affected by a variety of factors i.e. the melting temperature of the DNA hybrid ( $T_m$ ), the concentration of SSC in mol/l, the base composition % (G+C), the length of the shortest chain in the duplex ( $n$  residues) and the concentration of helix destabilizing agents such as formamide. An empirical formula for DNA probes greater than 50 nucleotides in length was described by Meinkoth & Wahl (1984) as follows:

$$T_m = 81.5^\circ\text{C} + 16.6 \log M + 0.41\% (\text{G+C}) - (500/n) - 0.61 (\% \text{ formamide})$$

Because the DNA probe had been nick-translated (Section 2.2.7.7) it was assumed that the shortest chain in the duplex ( $n$ ) was 100 residues. Empirical determination of  $T_m$  at a given molarity of SSC enabled Marmur & Doty (1962) to calculate the mole percent (%) G+C for a

whole variety of organisms including two plant species. Tobacco leaf DNA gave a  $T_m$  of 85.5°C and wheat germ DNA a  $T_m$  of 88.5°C, from which the mole percent G-C was calculated as 39% and 46% respectively. The latter value [46% (G+C) for wheat germ DNA] was used in conjunction with the equation described above to calculate the stringency of the post-hybridizational washes and is listed in Table 1. Bonner *et al.* (1973) reported that the melting temperature ( $T_m$ ) decreased 1°C for every 1% of mismatched base pair and this value was used to calculate the % mismatch ( $T_m - T$ ) for filters washed at a given stringency (Table 1).

Hybridization was carried out under conditions of relative low stringency (55°C and 5x SSPE) which allowed a high rate of hybridization, followed by a series of post-hybridization washes of increasing stringency. The filter was prehybridized in a sealed bag at 55°C for 3 h in preheated freshly made hybridization buffer (150  $\mu\text{l}/\text{cm}^2$  of filter) consisting of 5x SSPE (diluted from a 10x SSPE stock), 5x Denhardt's solution (diluted from a 100x stock solution) 0.2% (w/v) SDS (diluted from a 10% (w/v) SDS solution) 20  $\mu\text{g}/\text{ml}$  polyU (diluted from a 20 mg/ml stock solution in TE buffer), and 100  $\mu\text{g}/\text{ml}$  salmon sperm DNA (diluted from a stock solution of 10 mg/ml in distilled water). The 10x SSPE solution consisted of 200 mM  $\text{NaH}_2\text{PO}_4$ , 20 mM EDTA, pH 7.4. The 100x Denhardt's solution was prepared by dissolving Ficoll type 400, polyvinyl pyrrolidone (PVP) and BSA at a final concentration of 2% (w/v) in distilled water. The salmon sperm DNA stock solution (10 mg/ml) was prepared as described by Maniatis *et al.* (1982) and boiled freshly for 10 min to denature the dsDNA before used for dilution. Following the prehybridization the filter was transferred to a fresh plastic bag containing hybridization buffer (100  $\mu\text{l}/\text{cm}^2$  of filter) and a typical nick-translated DNA fragment (Section 2.2.7.7), freshly boiled for 10 min, containing about  $10^7$  cpm/ $\mu\text{g}$  DNA. The concentration of nick translated DNA fragment did not exceed 50-100 ng/ml of hybridization buffer and the hybridization was performed at 55°C for 20-24 h. The stringency at which the filter was washed depended on the % mismatch expected between the DNA fragment and the DNA bound to the nitrocellulose filter. The conditions are discussed in Section (3.1.4.3) with reference to Table 1.

Table 1. DNA hybridization parameters

Wash buffer	stringency	T <sub>m</sub> (°C)	% mismatch of base pairs	
			60°C wash	65°C wash
5 x SSC	low	94.0	34.0	24.0
2 x SSC		87.4	27.4	22.4
1 x SSC		82.3	22.3	17.3
0.5 x SSC		77.4	17.4	12.4
0.1 x SSC	high	65.8	5.8	< 1

The wash buffer also contained 0.1% (w/v) SDS

#### 2.2.11 DNA SEQUENCE DETERMINATION

The method used to determine the nucleotide sequence of a given piece of DNA was based on the one described by Murphy & Kanvanagh (1988) and required highly purified plasmid DNA in large quantities. An aliquot (20  $\mu$ l) of DNA isolated as described in Section 2.2.5.3.2, or highly purified DNA (10  $\mu$ g/20 $\mu$ l in TE<sub>0.1</sub> buffer) was denatured by addition of 1  $\mu$ l of 5 M NaOH, for 5-10 min at room temperature. In the meantime the minigel-filtration columns for the spindialysis were prepared as follows. The base of a small Eppendorf tube (0.8 ml) was punctured using just the tip of a syringe needle (23-Gauge x 1 $\frac{1}{4}$ ) and the base of a large Eppendorf tube (1.5 ml) was punctured using the same needle but pushing it fully through. The lids of both the tubes were cut off and the small tube was placed inside the large tube which was placed in a 5 ml plastic test tube which fitted the Rodwell H-103N centrifuge. Glass beads (Ballotini no. 1), which had been acid-washed and were stored at room temperature in TE<sub>0.1</sub> buffer, were transferred to the small tube to a depth of 2 mm after which the tube was filled to the rim with Sepharose CL-6B which had been prepared as a 66% (v/v) slurry in TE<sub>0.1</sub> buffer, autoclaved and stored at 4°C, using a sterile Pasteur pipette. The 5 ml plastic test tube containing the large and small Eppendorf tubes was placed in the Rodwell H-103N centrifuge and spun at 3400 g for 3 min at room temperature. The denatured DNA was then loaded on the column, the large Eppendorf tube replaced by a fresh one without a hole and the whole apparatus centrifuged again exactly as described above. Between 16-20  $\mu$ l of denatured DNA was recovered, and either used directly for determination of the nucleotide sequence or frozen and stored at -20°C.

An aliquot of the denatured DNA (8  $\mu$ l) was transferred to a fresh small Eppendorf tube, and 1  $\mu$ l of TM consisting of 100 mM TrisCl, 100 mM MgCl<sub>2</sub>, pH 8.0 and 1.5  $\mu$ l of primer at a concentration of 10  $\mu$ g/ml in distilled water, were added. If more than two DNA samples were used, a TM-mix consisting of TM and primer in a ratio of 1:15 was freshly prepared and an equal volume of 2.5  $\mu$ l of the TM-mix was added to each DNA sample. Annealing of the primer was established by incubation at 37°C for 15-20 min. In the meantime an aliquot

(2  $\mu$ l) of each of the dideoxy-mixes, A, T, C and G, which had been prepared as outlined in Table 2 and stored at  $-20^{\circ}\text{C}$ , was pipetted into a large Eppendorf tube and kept in ice-water. The Klenow-mix was prepared as follows (for 10 DNA samples): 9  $\mu$ l of TM, 62  $\mu$ l of distilled water, and 15  $\mu$ l of a 8,3 mM [ $\alpha$ - $^{35}\text{S}$ ]dATP solution (specific activity  $\pm 44.4 \text{ TBq/mmol} = 1200 \text{ Ci/mmol}$ ) to which 5  $\mu$ l Klenow fragment of DNA polymerase I (5 U/ $\mu$ l) was added were mixed together and kept in ice-water. An aliquot of the Klenow-mix (8  $\mu$ l) was added to each of the annealing reaction mixtures which had been allowed to cool to  $0^{\circ}\text{C}$  in ice-water. An aliquot of the Klenow-mix/annealing reaction (4.5  $\mu$ l) was placed on the inside of the wall of each Eppendorf tube containing one of the four dideoxy mixes. The reaction was started by a 2 s spin in an Eppendorf microfuge at room temperature, and continued by incubation at  $42^{\circ}\text{C}$  for 10 min.

After 8 min, an aliquot (2  $\mu$ l) of NTP-mix of 0.5 mM dNTPs (A, C, G and T) prepared and stored as described by Maniatis *et al.* (1982), was placed on the inside of wall of the Eppendorf tube. After 10 min the NTP-mix was spun to the bottom of the tube by means of an Eppendorf microfuge for 2 s at room temperature, and the incubation continued for an additional 5 min at  $42^{\circ}\text{C}$ . An aliquot (4  $\mu$ l) of formamide mix consisting of 10 ml deionized formamide (formamide stirred with BDH 'Amberlite' monobed resin MB-3 until the pH was 7.5, vacuum-filtered through Whatman no. 1 filterpaper in a Buchner funnel to remove the resin and aliquoted into 500  $\mu$ l portions which were stored at  $-20^{\circ}\text{C}$ ), 0.01 g bromophenolblue, 0.01 g xylene cyanol FF and 2 ml 0.5 M EDTA pH 8.0, was added, and the samples either boiled for 2 min or incubated in a heating-block at  $85^{\circ}\text{C}$  for 15 min; the latter procedure was used to concentrate the samples. An aliquot of the sample (5  $\mu$ l) was loaded onto a polyacrylamide sequencing gel which was run and carried through the procedure described in Section 2.2.12.1.3. After drying the gel, a sensitive monitor was used to determine the distribution of the radio activity across the gel. If the top of the gel registered 300-500 cps, an overnight exposure to X-ray film produced a strong signal.

Table2. Composition of dideoxynucleotide mixes\*

Solution**	A	C	G	T
0.5 mM d-CTP	135 $\mu$ l	10 $\mu$ l	200 $\mu$ l	200 $\mu$ l
0.5 mM d-GTP	135 $\mu$ l	200 $\mu$ l	10 $\mu$ l	200 $\mu$ l
0.5 mM d-TTP	135 $\mu$ l	200 $\mu$ l	200 $\mu$ l	10 $\mu$ l
10 mM dd-ATP	2 $\mu$ l	-	-	-
10 mM dd-CTP	-	8 $\mu$ l	-	-
10 mM dd-GTP	-	-	16 $\mu$ l	-
10 mM dd-TTP	-	-	-	53 $\mu$ l
TE <sub>0.1</sub> buffer	397 $\mu$ l	392 $\mu$ l	384 $\mu$ l	347 $\mu$ l

\* Mixes were aliquoted (22  $\mu$ l) and stored at -20°C.

\*\* The solutions were prepared in TE<sub>0.1</sub> buffer and stored at -20°C.

## 2.2.12 GELELECTROPHORESIS

### 2.2.12.1 Nucleic acid analysis

#### 2.2.12.1.1 DNA analysis on high resolution agarose gels

When a high resolution of DNA fragments was required agarose gels were made and run in TAE buffer consisting of 40 mM Tris-acetate, 2 mM EDTA pH 8.0, diluted from a 50x stock made up according to Maniatis *et al.* (1982); 50x TAE consisted of 242 g Tris, 57.1 ml glacial acetic acid and 100 ml 0.5 M EDTA, pH 8.0. The agarose concentration was dependent on the size of the DNA fragments to be resolved and varied from 0.9 - 1.2% (w/v) as described by Maniatis *et al.* (1982). The appropriate amount of agarose, medium EEO, type II was dissolved in 300 ml of TAE buffer using a microwave oven at low power for 5 min. The solution was cooled to approximately 60°C and poured into a gel mould with a width of 14 cm and a length of 15 cm. The type of comb used depended on the volume and the number of samples to be analyzed. After pouring, the gel was left to set at room temperature for at least 1 h. The gel was transferred to a Leicester Biocentre gel tank and submerged in TAE buffer, approximately 2 l. To prepare the DNA samples for analysis, a one-sixth volume of DNA loading buffer consisting of 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 15% (v/v) Ficoll type 400 in H<sub>2</sub>O, was added and the samples mixed. After loading of the samples on the gel, the DNA was subjected to electrophoresis for 16 h at 35V.  $\lambda$  phage DNA was digested with *Hind*III and *Eco*RI (Section 2.2.7.1) and always run on the same gel. After electrophoresis, the gel was stained in ethidium bromide (0.5  $\mu$ g/ml) for 45 min, on a shaker, destained in distilled water (3 washes) and the DNA visualized under u.v. light. A Polaroid Instant camera holding a Polaroid P665 or P667 film was used to photograph the gel.

#### 2.2.12.1.2 Rapid DNA analysis on agarose gels

For rapid analysis of digested plasmid DNA (Section 2.2.7.1) or M13 ssDNA (single-stranded DNA), prepared as described in Section 2.2.4.2, a small (40 ml) typical 1% (w/v) agarose gel (size 5.9 x 8.3 cm) was prepared in TBE buffer diluted freshly from a 10x stock solution consisting of 324 g Tris, 55 g boric acid, 19 g Na EDTA and distilled water to 1 l. After the agarose solution had cooled to approximately 60°C, ethidium bromide was added from a stock solution (10 mg/ml) to a final concentration of 5 µg/ml. The agarose was left at room temperature to set for at least 30 min, and the gel was submerged in TBE buffer in a BRL Horizon 58 gel tank. Plasmid DNA samples were mixed with one-sixth volume of DNA loading buffer. When M13 ssDNA was analyzed, 20 µl of the phage stock (approximately 10<sup>10</sup> pfu/ml) was added to 4 µl M13 loading dye consisting of 0.15% (w/v) bromophenol blue, 0.5% (w/v) SDS, 0.15 M EDTA pH 8, 50% (v/v) glycerol, which had been heated for 10 min at 65°C to dissolve the SDS. The M13 loading dye was stored at 4°C. The DNA samples were loaded and the gel was run for 30-60 min at 70-100 mA (Voltage ≤ 100 V). The DNA was visualized under u.v. light and when required a photograph taken as described in Section 2.2.12.1.1.

#### 2.2.12.1.3 DNA analysis under denaturing conditions

The complementary DNA strands synthesized *in vitro* during the sequence reactions (Section 2.2.11), were denatured and separated by the difference in their chain length on polyacrylamide gels as described by Bankier & Barrell (1984) and G. Murphy personal communication. A potential difference gradient was generated in order to increase the size range of fragments that could be separated as described by Biggin *et al.* (1983).

Electrophoresis was routinely performed using a 0.3 mm thick gel, casted between glass plates (20 cm x 50 cm). One glass plate, with a 2 cm x 17 cm notch was coated in  $\gamma$ -methacryloxypropyltrimethoxysilane (Ansorge & de Mayer, 1980), prior to assembly, to prevent the gel from sticking to this plate upon dismantling after electrophoresis. The plates



were washed thoroughly with detergent, rinsed with distilled water followed by ethanol, and dried. The glass plates were assembled by placing a 0.5 cm x 50 cm strip of 0.3 mm modeller's Plasticard along both 50 cm edges and 0.5 x 22 cm strip along the bottom edge between the plates, and gently clamping with fold back-clips.

The gradient gel consisted of a 1x to 5x TBE buffer gradient over approximately the bottom one third of the gel. Since the gradient did not need to be linear a rapid method of generating a rough gradient in a pipette was used. An aliquot (14  $\mu$ l) of 25% (w/v) aminonium persulphate (APS), freshly dissolved, and 7  $\mu$ l of N,N,N',N'-tetramethylenediamine (TEMED) was added to 7 ml of 5x TBE mix consisting of 4.6 g urea, 1.5 ml 40% (w/v) acrylamide solution (38% (w/v) acrylamide, 2% (w/v) bisacrylamide deionised with 20 g amberlite MB-1, and filtered through 2 layers of Whatman no. 1 filterpaper, stored in the dark at 4°C), 5 ml 10x TBE buffer (Section 2.2.12.1.2) and 200  $\mu$ l 2% (w/v) bromophenol blue, made freshly. To 35 ml of 1x TBE mix consisting of 60 ml 40% (w/v) acrylamide solution, 40 ml 10x TBE buffer, 184 g urea and distilled water to 400 ml, which was stored in the dark at 4°C, 70  $\mu$ l of 25% (w/v) APS plus 70  $\mu$ l of TEMED were added. First 6 ml of the 1x TBE mix and then 7 ml of the 5x TBE mix were drawn into a 10 ml pipette using a Pipetman. A rough gradient was generated by the careful introduction of 2-3 air bubbles using the Pipetman. The gradient solution was poured slowly into the gel template which was held at an angle of about 45° in a continuous flow down one edge. The gel plate was lowered to the bench as the last of the mix was poured. A 25 ml pipette was filled with 1x TBE mix and pouring recommenced. As the gel template filled to the top it was lowered onto the bench, filled to overflowing, and a slot former (44 slots) inserted so that about 3-5 mm of the comb above the slots entered the gel. The gel was left to set for 30 min and used within the next hour.

The gel was mounted onto a Raven vertical slabgel apparatus together with a cooling plate. The lower buffer chamber was filled with 5x TBE buffer and the top buffer chamber with 1x TBE buffer. The slot former was carefully removed and the wells rinsed several times with 1x TBE buffer. An aliquot (1-1.5  $\mu$ l) of denatured DNA samples (Section 2.2.11), were loaded, using a 20  $\mu$ l Gilson pipette and an applicator made from a Gilson tip

as follows. The end of a yellow tip was heated in a low bunsen burner flame until it softened. A pair of tweezers was used to hold the softened end while the rest of the tip was left to fall to the floor. After cooling the tip was trimmed and tested for its ability to fit between the gel plates.

The gel was run at a constant power of 40 Watt as this maintains a hot, denaturing environment in the gel. At this setting the voltage varies from 1.2-1.5 kV and the current between 20-28 mA. The gel was run for 3-3<sup>1</sup>/<sub>2</sub> hrs until the dye reached the bottom of the gel. The notched plate was prized off, using a thin spatula, and the gel on the bottom plate fixed in a solution consisting of 10% (v/v) methanol and 10% (v/v) acetic acid (2 l) for 25 min. The plate was agitated occasionally to help the gel float off the plate. The plate with the gel was removed from the fixing solution, excess liquid drained off and the gel transferred to Whatman 3 MM filterpaper by placing a sheet onto the gel and applying gentle pressure. The gel stuck to the paper and was peeled off the plate, after which a layer of Cling Film was used to cover the gel. The gel was placed on a geldryer for 20 min at 80°C. The Cling film was removed and the dried gel exposed to an X-ray film placed inside a film cassette. Using this protocol (see also Section 2.2.11) an overnight exposure at room temperature was sufficient to yield a useful image.

#### 2.2.12.1.4 RNA analysis on gels containing formaldehyde

RNA transcripts produced from SP<sub>6</sub>/T<sub>7</sub> transcription vectors (Section 2.2.17.1) were analyzed under denaturing conditions on agarose gels containing formaldehyde, according to a method modified from that described by Lehrach *et al.* (1977). The concentration of agarose used depended on the size of the transcript to be analyzed. A transcript of about 2 kb was analyzed on a 1.5% (w/v) agarose gel. To 2.25 g of agarose 112.5 ml distilled water and 15 ml 10x Mops buffer consisting of 0.2 M 3-(N-morpholino)propanesulfonic acid (Mops) pH 7.0, 50 mM Na acetate and 10 mM EDTA, was added. The agarose was dissolved using a microwave oven at low power for 5 min and the solution allowed to cool to 60°C. An aliquot (22.5 ml) of filtered formaldehyde (a commercial formaldehyde solution

40% (w/v), filtered through Whatman no.1 paper to remove any paraformaldehyde that might have precipitated out of solution) was added, mixed and the gel (14 cm x 15 cm) was poured in the fumehood. The gel was left to set for at least 1 h, after which it was submerged in 1 x Mops buffer.

An appropriate amount of RNA, which was up to 10 µg in 5 µl distilled sterile water, was mixed with 15 µl denaturing solution consisting of 500 µl of deionised formamide (see Section 2.2.11), 100 µl 10x Mops buffer and 150 µl of filtered formaldehyde. The sample was heated at 60°C for 5 min, allowed to cool to room temperature and 2 µl RNA loading buffer consisting of 50% (v/v) glycerol, 0.2% (w/v) bromophenol blue, added. The wells were flushed with 1x MOPS buffer to remove the formaldehyde and the samples loaded immediately using a 20 µl Gilson pipette. The gel was run for 3 h at 100 V (approximately 80 mA) until the dye was two-thirds the way down the gel, soaked in 250 ml 10% (w/v) glycine for 20 min with shaking, stained with ethidium bromide, by adding 100 µl of 10 mg/ml, and destained with two changes of distilled water for 15 min each. The RNA and the DNA markers (phage λ DNA digested with an appropriate restriction enzyme) were visualized under u.v. light and a photograph taken. When the transcript and/or the DNA had been labelled with <sup>32</sup>P the gel was vacuum dried on a gel dryer at 80°C for about 1 1/2 h and placed against X-ray film for autoradiography.

#### 2.2.12.2 Protein analysis

##### 2.2.12.2.1 Protein analysis under non-denaturing conditons

Analysis of native proteins was performed on polyacrylamide gels using a modification of the method described by Hedrick & Smith (1968). In order to increase the size range of proteins to be separated a linear acrylamide gradient was generated, either 4-15% (w/v) or 4-30% (w/v) polyacrylamide. The different acrylamide mixes were prepared as listed in Table 3. A gel cassette was assembled from two glass plates, either 16.5 cm square for large gels on 13 cm x 11.5 cm for small gels and two side spacers, 0.1 cm in thickness. Silican tubing

Table 3. Composition of non-denaturing protein gel mixes

Solution	Gel mixes				
	large gel (16.5 x 16.5 cm)			small gel (13 x 11.5 cm)	
	4%	15%	30%	4%	15%
60% (w/v) acrylamide 0.3% (w/v) bisacrylamide*	-	7.2 ml	14.4 ml	-	2.5 ml
25% (w/v) acrylamide 1% (w/v) bisacrylamide*	4.6 ml	-	-	1.6 ml	-
30% (v/v) glycerol	-	10.6 ml	10.6 ml	-	3.7 ml
3M Tris HCl pH 8.8	3.75 ml	3.75ml	3.75 ml	1.3 ml	1.3 ml
Distilled water	20.4 ml	7.2 ml	-	7.0 ml	2.4 ml
10% (w/v) APS**	150 $\mu$ l	75 $\mu$ l	75 $\mu$ l	50 $\mu$ l	25 $\mu$ l
TEMED	18 $\mu$ l	9 $\mu$ l	9 $\mu$ l	6 $\mu$ l	3 $\mu$ l

\* Acrylamide solutions were stored at 4°C in the dark.

\*\* Solution was freshly made, kept at room temperature and never used after five days.

was used to seal the bottom and the sides, after which the plates were clamped with fold-back clips. An equal volume of two acrylamide mixes i.e. either 12.5 ml (large gels) or 7.6 ml (small gels), was pipetted into each chamber of a MSE gradient mixer, which holds a volume of 2 x 30 ml for large gels and 2 x 10 ml for small gels. The gradient mixture was used in connection with an LKB peristaltic pump to produce the gradient. The speed of the peristaltic pump was such that it took approximately 5 min to pour the gradient gel. A slot former (12 or 14 slots) was inserted into the gel solution, which was left to polymerise for at least 1 h at room temperature.

The gel was transferred to a Raven vertical slabgel apparatus and the top tank filled with freshly made running buffer consisting of 50 mM Tris-HCl, 384 mM glycine pH 8.5 containing 8 mM L-cysteine, while the bottom tank was filled with running buffer from which L-cysteine was omitted. In order to ensure reducing conditions in the gel a current of 15 mA was passed for 1 h before the samples were loaded. The freshly prepared protein samples were mixed with loading medium consisting of 50% (v/v) glycerol, 0.01% (w/v) bromophenol blue to give a final concentration of 10% (w/v) glycerol and loaded on the gel using a 100 µl Hamilton syringe. The gel was run for 20 h at either 18 mA (large gel 4% - 30% (w/v) gradient), or 20 h at 8 mA (large gel 4% - 15% (w/v) gradient), or 6 h 20 mA (small gel 4% - 15% (w/v) gradient). Proteins in the gel were stained (Section 2.2.16) or immunoblotted (Section 2.2.14). Molecular weight markers used on these gels were urease hexamer (Mr 575 000), urease trimer (Mr 272 000), bovine serum albumin (BSA) dimer (Mr 132 000), BSA monomer (Mr 66 000), Ovalbumin (Mr 45 000), Carbonic anhydrase (Mr 29 000) and bovine lactalbumin (Mr 14 200). The gels were dried and stored at room temperature in a dry place.

#### 2.2.12.2.2 Protein analysis under denaturing conditions

Proteins were denatured in the presence of 2-mercaptoethanol and sodium dodecyl sulphate (SDS). Analysis on polyacrylamide gels containing SDS was performed according to the method initially described by Laemmli (1970). The gelplates were assembled as

described in Section 2.2.12.2.1 and the gel mix prepared as listed in Table 4. The percentage polyacrylamide chosen was dependent on the size range of polypeptides to be separated. The running mix was poured between the plates to within 2 cm from the top of the notched plate, and overlaid with butanol-1 saturated with distilled water, using a pasteur pipette. The gel was allowed to polymerize for 30-60 min. The butanol-1 and the water were poured off and the top of the gel washed several times with distilled water. The water left on top of the gel was removed using a piece of 3 MM Whatman paper. The stacking gel mix was poured on top, a slotformer (14 slots) inserted and the gel left to polymerize for approximately 10-15 min. The slot former was removed and the wells rinsed with distilled water, after which the gel was mounted onto a Raven vertical slab gel apparatus; the top and bottom compartment were filled with SDS-electrophoresis buffer consisting of 25 mM Tris Cl, 192 mM glycine 0.1% (w/v) SDS pH 8.3, freshly diluted from a 10x stock.

Protein samples were incubated with an equal volume of SDS loading buffer consisting of 12.5 mM Tris-HCl, 20% (v/v) glycerol, 4% (w/v) SDS, 10% (v/v) 2-mercaptoethanol, and 0.1% (w/v) bromophenol blue, for 5 min at 100°C. The samples were allowed to cool to room temperature and loaded on the gel by means of a 100 µl Hamilton syringe. The gel (13x11.6 cm) was run at 30 mA for 3-4 h until the dye had migrated to the bottom of the gel. Care was taken that the voltage did not exceed 200 V. Polypeptides in the gel were stained (Section 2.2.16) or immunoblotted (Section 2.2.14). Prestained molecular weight markers were prepared according to the manufacturers instructions. The mixture contained α2-macroglobulin (Mr 180 000), β-galactosidase (Mr 116 000), fructose-6-phosphokinase (Mr 84 000), pyruvate kinase (Mr 58 000), fumarase (Mr 48 500), lactic dehydrogenase (Mr 36 500) and triose phosphoisomerase (Mr 26 000). The gels were dried onto 3 MM Whatman paper on a vacuum gel dryer and stored in a dry place at room temperature.

Table 4. Composition of SDS denaturing protein gel mixes

Solution	Gel mixes (13 x 11.6 cm)			
	Running-mix			Stacking mix
	12.5%	15%	6%	10%
30% (w/v) acrylamide 0.3% (w/v) bisacrylamide*	6.6 ml	8.3 ml	10.0 ml	1.0 ml
1 M Tris-HCl pH 8.8	7.5 ml	7.5 ml	7.5 ml	-
0.5 M Tris-HCl pH 6.8	-	-	-	1.25 ml
Distilled water	5.7 ml	4.0 ml	2.25 ml	2.5 ml
10% (w/v) SDS	0.2 ml	0.2 ml	0.2 ml	0.05 ml
10% (w/v) APS	90 $\mu$ l	90 $\mu$ l	90 $\mu$ l	25 $\mu$ l
TEMED	9 $\mu$ l	9 $\mu$ l	9 $\mu$ l	4 $\mu$ l

\*Acrylamide solution was stored at 4°C in the dark.

### 2.2.13 PRODUCTION AND PURIFICATION OF ANTIBODIES

#### 2.2.13.1 Production of antibodies

New Zealand white female rabbits were used to raise antibodies against the *groEL* protein from *E.coli* and the Rubisco subunit binding protein from *Tritium aestivum*. The purified *groEL* protein was kindly provided by A Mehlet of the Hammersmith Hospital in London, and a sample of the purified Rubisco subunit binding protein was a generous gift of A.J. Keys from the Rothamsted Experimental Station in Harpenden. Both protein samples were lyophilized. The immunization of the rabbits was basically carried out as described by Johnstone & Trope (1982).

A blood sample of 5 ml was taken from each rabbit prior to immunization and serum prepared as follows. The blood was removed from the rabbit's ear, collected in a 25 ml Universal vial, and a Pasteur pipette used to streak along the sides of the vial in order to prevent clotted blood sticking to the wall. The blood was allowed to clot at room temperature. The serum was isolated by centrifugation in a MSE Coolspin centrifuge at 3800 g for 20 min at 15°C and the supernatant carefully transferred to a fresh Universal tube, mixed and aliquots of 200-500 µl were frozen and stored at -20°C. This pre-immune serum was analyzed for antibody activity against specific proteins in total protein extracts of both *E.coli* and wheat leaves, which had been analyzed by SDS polyarylamide gelelectrophoresis and transferred to nitrocellulose (Section 2.2.14.1). Only if the result of this experiment was negative i.e. no cross-reactive antibodies were present in the pre-immune serum, were the rabbits used for immunization, which was carried out as follows.

The rabbit was injected subcutaneously with 100 µg of protein dissolved in 200 µl phosphate buffer saline (PBS) consisting of 8 g NaCl, 0.2 g KCl, 11.5 g Na<sub>2</sub>HPO<sub>4</sub>, 0.2 g KH<sub>2</sub>PO<sub>4</sub>, distilled water to 1 l, pH 7.4, which had been mixed with 200-300 µl Freund's incomplete adjuvant (Gibco Ltd.). After 2 weeks a second injection of 100 µg protein was given as described above. A blood sample (10 ml) was taken from the rabbit's ear after 7 days and serum prepared and analyzed as described above. When a strong antibody response



was observed, a second blood sample was taken, 12 days after the second injection. In case there was a low or no antibody response the rabbit was injected again with 100 µg protein 20 days after the second immunization, a blood sample was taken 7 days after the third injection and serum prepared and analyzed as described above.

#### 2.2.13.2 Purification of antibodies

The standard procedure for raising antibodies involves the addition of Freund's complete adjuvants to the antigen solution before injection Johnstone & Trope (1982). Freund's complete adjuvants is a mixture of oil (Bayol F) and detergent (Mannide mono-oleate) containing dead *Mycobacterium tuberculosis* cells (Gibco Ltd, personal communication), the latter being added to boost the immune system of the rabbit upon injection. This procedure however results in the rabbit also producing antibodies against mycobacterial proteins. These antibodies may interfere with the detection of specific antigens in plant protein extracts if these antigens are related to mycobacterial proteins. Steps were taken to remove those antibodies from the serum, which recognize bacterial proteins.

##### 2.2.13.2.1 Purification of antibodies using a λ gt11 plate lysate

In order to isolate specific cDNA sequences from a λ gt11 expression library (Section 2.2.6) it is essential to remove those antibodies, which recognize host *E.coli* and λ gt11 phage proteins, from the serum. A plate lysate of non-recombinant λ gt11 was prepared by method 1 as described by Maniatis *et al.* (1982), using four 85 mm Petri dishes. A nitrocellulose filter with a diameter of 150 mm was soaked in PBS and incubated with the plate lysate in a sealed bag for 2 h at 30°C in a shaker. The lysate was removed, Na azide added to a final concentration of 0.02% (w/v), and the mixture stored at 4°C so that it could be used again. The nitrocellulose filter was incubated for an additional 2 h in PBS containing 3% (w/v) Marvel and 0.02% (w/v) Na azide at room temperature with shaking. The antibody solution (i.e. serum) was diluted 50 to 100 fold depending on the titre of the serum, in PBS

containing 2% (w/v) BSA, and incubated with the nitrocellulose filter, which had been rinsed in PBS, in a sealed bag at 30°C overnight with shaking. The antibody solution was drained off, Na azide added to a final concentration of 0.02% (w/v), and aliquots frozen and stored at -20°C. The purified antibody solution was analyzed for its ability to cross-react with the antigen of interest but not to recognize either any host *E.coli* or  $\lambda$  phage proteins by means of methods described in Section 2.2.6, 2.2.12.2.2 and Section..

#### 2.2.13.2.2 Purification of antibodies using a total *E.coli* protein extract.

The procedure used followed the one described in Section 2.2.13.2.1, with the difference that instead of an *E.coli* plate lysate, an *E.coli* total protein extract was used. In order to prepare such an extract, 10 ml of LB medium was inoculated with a single *E.coli* colony from a fresh LB agar plate and incubated at 37°C overnight. LB medium (200 ml) was inoculated with 2 ml of the overnight culture, and the cells were allowed to grow at 37°C with shaking until the density was such that the absorbance (A) at 600 nm was 1-1.2. The cells were harvested by centrifugation in a MSE Coolspin centrifuge at 3800 g for 20 min at 4°C and resuspended in sonication buffer, consisting of 100 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 5 mM EDTA, 1 mM 2-mercaptoethanol, 1 mM phenylmethylsulphonyl fluoride (PMSF), pH 8.0, at a concentration of about 1 g wet weight *E.coli* cells plus 5 ml of sonication buffer. Cells were sonicated four times for 30 s each, at 0°C, after which clearing of the suspension was observed. Soluble proteins were separated from the cell walls and membranes by centrifugation in a MSE18 centrifuge at 10 000 g for 15 min at 4°C. The supernatant was incubated with a nitrocellulose filter which was been presoaked in PBS, and the procedure continued as described in Section 2.2.13.2.1.

## 2.2.14 IMMUNOBLOTTING OF PROTEINS AND ANTIBODY DETECTION

### 2.2.14.1 Immunoblotting of proteins

Protein samples to be analyzed by immunoblotting were first subjected to polyacrylamide electrophoresis under either non-denaturing (Section 2.2.12.2.1) or denaturing (Section 2.2.12.2.2) conditions. The method used was a modification of the one described Renart *et al.* (1979) and by Towbin *et al.* (1979). After electrophoresis, the polyacrylamide gel was washed in transfer buffer consisting of 192 mM glycine, 25 mM Tris, 20% (v/v) methanol, with constant shaking, for 30 min at room temperature. The proteins were transferred to nitrocellulose using a Biorad Transblot apparatus, which was run for 3 h at 150 mA at room temperature. When high molecular weight proteins were transferred i.e. from a gel run under non-denaturing conditions, 0.1% (w/v) SDS was dissolved in the transfer buffer. After transfer, the polypeptides remaining in the gel and if required those on the nitrocellulose were visualized (Section 2.2.16). The nitrocellulose filter was blocked by washing in PBS containing 1% (w/v) Marvel for at least 30 min at room temperature under constant shaking. The filter was rinsed with PBS twice, and was either dried and stored at room temperature if required, or immediately used for incubation with serum containing antibodies raised against a particular protein of interest (see also Section 2.2.13.2.2). The serum was normally used in a 1:1000 to 1:5000 dilution in PBS. The filter, 10 cm square, was incubated with 5 ml of diluted antibody solution in a sealed bag at 30°C overnight with shaking. The filter was rinsed in PBS twice, and the antibody detected using one of the methods described in Section 2.2.14.2.

#### 2.2.14.2 Antibody detection

##### 2.2.14.2.1 Enzyme-linked detection

The method used was a modification of the biotin-streptavidin system sold by Amersham International and described by the manufacturer. All the washes were performed on a shaker at room temperature. The filter was washed three times in PBS containing 0.1% (v/v) Tween 20 for 10 min per wash, after which it was placed in a plastic bag containing 5 ml PBS (for a maximum filtersize of 15 cm square) and 16.5  $\mu$ l biotinylated protein A (from Amersham International Plc.); the bag was sealed and incubated at 30°C for 60 min with shaking. The filter was washed in PBS containing 0.1% (v/v) Tween 20 with three changes of 10 min each, and transferred to a plastic bag containing 5 ml PBS plus 16.5  $\mu$ l streptavidin-linked peroxidase (from Amersham International Plc). After the bag was sealed, the filter was incubated at 30°C for 30 min with shaking, followed by two washes, in PBS containing 0.1% (v/v) Tween 20 for a total of 20 min, followed by two washes in PBS of 10 min each. A developing mix was prepared essentially described by Nakane (1968) by mixing solution A, consisting of 1.5 g NaCl, 1.0 ml 1 M Tris-HCl pH 7.5, and distilled water to 50 ml, to which 50  $\mu$ l 30% H<sub>2</sub>O<sub>2</sub> was freshly added, with solution B, consisting of 30 mg 4-chloro-1-naphthol dissolved in 10 ml methanol. The filter was developed by adding developing mix under constant shaking. Purple-blue bands appeared after approximately 5 min. If the reaction was slow, the filter was left to develop for 20 min, after which it was washed in distilled water (3 changes in a total of 30 min). The filter was dried between Whatman 3 MM paper overnight and stored at room temperature in a dry place.

#### 2.2.14.2.2 Radio-labelled detection

Special safety precautions were taken as this method involved the use of [ $^{125}\text{I}$ ] protein A, which was synthesized as described in Section 2.2.15. After immunoblotting, the filter was washed twice in PBS at room temperature on a shaker; incubation with the labelled protein A was carried out in a plastic bag which contained 5 ml of PBS and a volume of [ $^{125}\text{I}$ ] protein A solution equivalent to 500,000 cpm (Section 2.2.15) for a total of 2 h. The filter was washed twice in PBS containing 1% (v/v) Triton X-100 for 5 min each one final wash in PBS for 5 min, dried between Whatman 3 MM paper in a 60°C oven, and exposed to X-ray film using an intensifying screen.

#### 2.2.15 PREPARATION OF IODINATED PROTEIN A

To radiolabel protein A from *Staphylococcus aureus* the method outlined in the Pierce Chemical Company leaflet no 28666 was followed; this method is modified from the one initially described by Markwell (1982). All operations were carried out in a high efficiency fumehood in a radioisotope room. Two Iodo-beads were placed in a 25 ml plastic Universal vial and 0.5 ml PBS was added and left at room temperature for 5 min. The PBS was removed using a Pasteur pipette and the Iodo-beads washed twice again by the same procedure. Na[ $^{125}\text{I}$ ] (500 uCi), taken from a stock solution (specific activity 574 MBq/ $\mu\text{g}$  iodine = 15.5 mCi  $\mu\text{g}$  iodine), was diluted to 200  $\mu\text{l}$  with PBS, added to the washed Iodo-beads and left at room temperature for 5 min. The protein A suspension (5  $\mu\text{l}$  of 1 mg/ml in PBS), was added to the reaction mixture and left to stand at room temperature for 10 min. The reaction was stopped by removing the reaction mixture carefully with a Pasteur pipette. The iodinated protein A was separated from the free Na-iodine by gel filtration on a Sephadex G-100 superfine column in PBS.

The column was made in a disposable 3-5 ml plastic pipette. The protein A was eluted from the column with PBS and 0.5 ml fractions were collected in Eppendorf tubes. An aliquot (2  $\mu\text{l}$ ) of each fraction was added to 1 ml of scintillation fluid (LKB 'Optiphase'

safe) and counted in a scintillation counter (LKB 1280 ultragamma). A typical iodination procedure gave approximately 370 000 cpm/ $\mu$ l for fraction 3 and 160 000 cpm/ $\mu$ l for fraction 4. The iodinated protein A was stored at 4°C in a lead pot. A volume of iodinated protein A equivalent to 500 000 cpm was added to 5 ml of PBS and used to detect the antibodies bound to antigens after immunoblotting (Section 2.2.14.2.2).

#### 2.2.16 PROTEIN DETECTION

The method used to detect a protein depended on the physical state of the protein, i.e. whether it was bound to a filter, trapped in a gel or in aqueous solution. Proteins bound to a nitrocellulose filter were stained by soaking the filter in a solution consisting of 0.1% (w/v) amidoblack, 25% (v/v) propanol-2, 10% (v/v) acetic acid for 2 s, followed by immediate transfer to a destain solution which consisted of 25% (v/v) propanol-2, 10% (v/v) acetic acid. Destaining was performed under constant shaking in several (2-5) changes of destain solution until blue bands were visualized on a white background.

Proteins present in polyacrylamide gels were stained with Coomassie brilliant blue R250 by soaking the gel in stain solution, consisting of 0.5% (w/v) Coomassie brilliant blue R250, 40% (v/v) methanol, 7% (v/v) acetic acid, for 30 min at room temperature with shaking. Visualization of the protein band required removal of excess dye from the gel by diffusion into destaining solution, which consisted of 40% (v/v) methanol, 7% (v/v) acetic acid. The gel was submerged in the destaining solution and shaken gently with several changes (4-5) until the background was colourless. Radio-labelled proteins present in polyacrylamide gels were detected by fluorography, i.e. soaking the gel in Amplify (Amersham International) for 15 min at room temperature, drying of the gel under vacuum at 80°C for 1-2 h followed by autoradiography.

Proteins in solution were determined quantitatively by a modification of the method described by Bradford (1976) using the Bio-Rad protein assay kit. The standard assay procedure was followed as described by the manufacturer to determine the protein

concentration in the range of 200-1400  $\mu\text{g}$  of protein per ml of sample, and BSA was used as the protein standard.

#### 2.2.17 TRANSCRIPTION AND TRANSLATION *IN VITRO*

Specific DNA sequences of interest which were analyzed for their ability to be transcribed and translated *in vitro*, were cloned into the Riboprobe Gemini pGEM vector, sold by Promega Biotec Ltd, as first described by Yanisch-Perron *et al.* (1985).

##### 2.2.17.1 *Transcription in vitro*

The synthesis of RNA *in vitro* was carried out according to a modification of the method described by Krieg & Melton (1987). Plasmid DNA was linearized by digestion with an appropriate restriction enzyme (Section 2.2.7.1) for 16 h at 37°C. The DNA was extracted with phenol (Section 2.2.5.1.1), precipitated with ethanol (Section 2.2.5.1.2), and resuspended in 10 mM Tris-HCl pH 8.0. An aliquot (0.5  $\mu\text{g}$ ) was analyzed on an agarose gel (Section 2.2.12.1.2) to confirm linearization and recovery of the DNA. For the synthesis of RNA, approximately 1  $\mu\text{g}$  of linearized DNA was incubated in a total volume of 20  $\mu\text{l}$  of reaction-mix consisting of 2 mM spermidine, 40 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid (Hepes)-KOH, 6 mM Mg acetate, 10 mM DTT, 100  $\mu\text{g}/\text{ml}$  BSA, 0.5 mM of ATP, CTP and TTP, 0.1 mM GTP, 0.25 mM  $\text{m}^7\text{G}(5')\text{ppp}(5')\text{G}$ , 20 U RNase inhibitor (RNasin) and either 15 U SP6 RNA polymerase or 15 U T7 RNA polymerase, pH 7.5. The reaction mix, apart from the RNasin and RNA polymerases was a generous gift from M.J. May. The incubation was performed at 40°C for 30 min, at which time GTP in 20 mM Hepes-KOH pH 7.5 was added to a final concentration of 0.4 mM, and the incubation continued for a further 30 min. The transcriptional efficiency increased by approximately 10% if the RNA was not 'capped'. Uncapped RNA had been found to translate properly in a wheat germ translation system (M.J. May, personal communication),

and was synthesized by omitting  $m^7G(5')ppp(5')G$  from the reaction mix, and changing the incubation time to 1 h at 40°C.

Aliquots (1  $\mu$ l) from transcription reaction mixtures supplemented with an aliquot (1-10  $\mu$ Ci) of [ $^{32}P$ ] CTP (specific activity  $\pm 110$  TBq/mmol =  $\pm 3000$  Ci/mmol), were analyzed by electrophoresis (Section 2.2.12.1.4), followed by fluorography (Section 2.2.12.2.2), in order to determine whether a single major transcript of the expected size was generated. The efficiency of the reaction was calculated from the incorporation of [ $^{32}P$ ] CTP into the RNA. An aliquot (1  $\mu$ l) of transcription reaction mixture was used, following the method described in Section 2.2.7.7.

#### 2.2.17.2 Translation *in vitro*

The method used to synthesize polypeptides in a cell-free protein-synthesizing system from wheat germ was that described by Anderson *et al.* (1983). An aliquot (1-5  $\mu$ l) of a transcription reaction mixture (Section 2.2.17.1) was added to 3.75  $\mu$ l of wheat germ extract (prepared as described by Anderson *et al.*, 1983), 2.35  $\mu$ l 'energy-mix' consisting of 14 mM Hepes, 1 mM ATP, 8 mM creatine phosphate, 40  $\mu$ g/ml creatine- phosphokinase, 30  $\mu$ g/ml spermidine, 1.7 mM DTT, 20  $\mu$ M GTP, 25  $\mu$ M each of all the amino acids except methionine, 1 mM Mg (OAc) $_2$ , 84 mM KOAc, pH 7.6, and 1  $\mu$ l of a 15 nM L-[ $^{35}S$ ]methionine solution (specific activity  $> 37$  TBq/mmol =  $> 1000$  Ci/mmol) and distilled water to a final volume of 12.5  $\mu$ l. The wheat germ extract and the 'energy-mix' were a generous gift of J.E. Musgrove. The reaction mixture was incubated at 28°C for 1 h, after which the amount of L-[ $^{35}S$ ] methionine incorporated into the polypeptides was determined as follows.

An aliquot (2  $\mu$ l) of the reaction mixture was spotted onto the centre of a 2.4 cm disc of Whatman 3 MM filter paper and left to dry. The filter paper was boiled in a 10% (w/v) trichloroacetic acid (TCA) solution (10 ml) for approximately 2 min, allowed to cool to room temperature and washed once in 10% (w/v) TCA and twice in ethanol. The washed filter



paper was allowed to dry in air, 4 ml of liquid scintillation fluid (LKB, 'Optiphase' safe) was added and the radioactivity on the filter determined by means of a scintillation counter (LKB 1219, Rackbeta). An aliquot (3  $\mu$ l) of the translation mixture was analyzed by SDS polyacrylamide gelelectrophoresis and followed by fluorography (Section 2.2.12.2.2).

## 2.2.18 IMPORT OF POLYPEPTIDES BY ISOLATED INTACT CHLOROPLASTS

### 2.2.18.1 Isolation of chloroplasts

Chloroplasts were isolated from eight to nine-days-old *Pisum sativum* (pea) seedlings, following the method described by Ellis & Hartley (1982). The pea plants were grown as described in Section 2.2.1.1, except that the light intensity was reduced to 20  $\mu$ E,  $M^{-2}$ ,  $s^{-1}$ . This reduction was used in order to minimize the formation of starch grains which have been reported to affect the yield of intact chloroplasts (Ellis & Hartley, 1982).

The shoots of pea plants (25 g) were homogenised in 200 ml of a sterile frozen slurry of sucrose isolation medium (SIM) consisting of 0.35 M sucrose, 25 mM Hepes, 2 mM EDTA, pH 7.6, by means of a Polytron homogeniser (Northern Media Supply, Ltd.) for 6 s at 75% full speed. The homogenate was quickly filtered through eight layers of muslin into a pre-cooled beaker on ice and the filtrate centrifuged in a MSE Coolspin centrifuge at 3200 g for 1 min at 4°C. The pellet was resuspended using a cotton bud in 20 ml of SIM medium and the centrifugation repeated. The pellet was resuspended by means of a cotton bud in sorbitol reaction medium (SRM) consisting of 50 mM Hepes-KOH, 330 mM sorbitol, pH 8.4, to a final volume of 4 ml. The chloroplasts were further purified by Percoll gradient centrifugation following a modification of the method described by Morgenthaler *et al.* (1975).

A stepgradient of 20 ml was prepared containing 5 steps of 10% (v/v), 27.5% (v/v), 45% (v/v), 62.5% (v/v) and 80% (v/v) Percoll in SRM, and kept on ice. An aliquot (2 ml) of the chloroplast suspension was loaded onto the gradient and centrifuged in a MSE Mistral 6L centrifuge at 3200 g for 15 min at 4°C. The rotor was allowed to come to a rest without the

aid of the brake. The intact chloroplasts, present in the 45% (v/v) Percoll step, were removed from the gradient using a Gilson 1 ml pipette holding a tip with the end cut off. The chloroplasts were washed twice by adding 4 ml of SRM followed by centrifugation in a MSE Coolspin centrifuge at 4000 g for 2 min at 4°C. The final pellet, containing the purified chloroplasts, was resuspended in SRM medium at a chlorophyll concentration of about 1 mg/ml.

The chlorophyll concentration was determined by the method of Arnon (1949). An aliquot of the chloroplast suspension (20-50 µl) was diluted with SRM and acetone to a final concentration of 80% (v/v) acetone and incubated in the dark at room temperature for 10 min. The protein was sedimented by centrifugation in an Eppendorf microfuge at 11,600 g for 5 min at room temperature. The absorbance (A) of the supernatant was measured at 645 nm and 663 nm and the chlorophyll concentration calculated using the following formula:

$$\mu\text{g ml}^{-1} \text{ chlorophyll} = (20.2 \times A_{645}) + (8.02 \times A_{663})$$

The chloroplast suspension was used to import polypeptides, synthesized *in vitro* (Section 2.2.18.2).

#### 2.2.18.2 Import of polypeptides by chloroplasts

The procedure used to import synthesized polypeptides *in vitro* by intact isolated chloroplasts was based on the method described by Robinson & Ellis (1985). An aliquot (100 µl) of chloroplast suspension (Section 2.2.18.1), 20 µl of SRM medium, 3 µl of 5x SRM medium plus an aliquot (10 µl) of a desired translation mixture, containing the polypeptide of interest (Section 2.2.17.2), were mixed and kept on ice. Incubation was carried out under illumination (100 µE, m<sup>-2</sup>, s<sup>-1</sup>) at 25°C with occasional gentle shaking. After the required incubation time, the sample was transferred to ice and 6 ml of cold SRM medium added, followed by centrifugation in a MSE Coolspin centrifuge at 4000 g for 2 min at 4°C. The chloroplasts in the pellet were lysed by adding 100 µl 10 mM Tris-HCl pH 8.0

to the pellet, and the chloroplast membranes separated from the stromal fraction by centrifugation in an Eppendorf microfuge at 11,600 g for 10 min at 4°C. The supernatant was analyzed by polyacrylamide gel electrophoresis either under non-denaturing or denaturing conditions as described in Section 2.2.12.2.1 and Section 2.2.12.2.2 respectively.

#### 2.2.19 SYNTHESIS OF FOREIGN PROTEINS IN *ESCHERICHIA COLL*

The method used to synthesize foreign proteins in *E.coli* was that described by van der Vies *et al.* (1986). The cDNA sequences, encoding a polypeptide of interest, were cloned into expression vectors i.e. either a plasmid or a M13 phage vector; the different constructs are described in detail in Section 3.4 and 3.5.

The different plasmids are able to replicate in *E.coli* strain TG<sub>2</sub> and contain the cDNA sequences under the transcriptional control of the *lacZ* promoter. An aliquot (10 ml) of YT medium (Section 2.2.2) with the appropriate antibiotic was inoculated with a single colony of TG<sub>2</sub> *E.coli* cells, sometimes containing a plasmid, from a fresh YT plate (Section 2.2.2.1) and incubated overnight at 37°C. The overnight culture was diluted 100-fold in fresh YT medium containing the required antibiotic and 0.5 mM IPTG, the latter to induce transcription. The culture was incubated at 37°C with constant orbital shaking until the absorbance (A) at 600 nm reached 1.2-1.4. The cells were harvested by centrifugation in a MSE 18 centrifuge at 4000 g for 15 min at 4°C. The wet weight of the cells was determined and the cells were either frozen at -20°C or used immediately to prepare a 'soluble' protein extract (Section 2.2.20.1).

TG<sub>2</sub> cells, which sometimes contained a plasmid, and were to be used for infection by M13 phage, were grown in 10 ml of MG medium containing the appropriate antibiotic from a single colony taken from a MG plate (Section 2.2.2) for 30 h at 37°C in order to maintain selection for F'. The overnight culture was diluted 100-fold in fresh YT medium containing antibiotic and 0.5 mM IPTG, and incubated until the absorbance (A) of the cell culture at 600 nm was 0.8 - 1.0. The cells were infected with an aliquot of the desired M13 phage stock

solution at a final concentration of  $10^{10}$ - $10^{11}$  pfu/ml cellculture, and grown for an additional 2 h before harvesting, weighing and storage as described above.

#### 2.2.20 PREPARATION OF A PROTEIN EXTRACT

##### 2.2.20.1 Preparation of a 'soluble' *Escherichia coli* protein extract

To prepare a 'soluble' protein extract the *E.coli* cells were resuspended in isolation buffer consisting of 100 mM Tris-HCl, 20 mM  $MgCl_2$ , 10 mM  $NaHCO_3$ , 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 8.0, at a concentration of approximately 1 g of wet weight *E.coli* cells plus 1 ml of isolation buffer. Alterations in the isolation buffer as described here are indicated in the specific sections. The cell suspension was sonicated four times for 30 S each at 0°C, after which clearing of the suspension was observed, and centrifuged in a Beckman TL100 centrifuge at 100 000 g for 30 min at 4°C. The supernatant was carefully transferred to a fresh Eppendorf tube by means of a Gilson pipette, avoiding the interface which was sometimes observed due to the high protein concentration; the supernatant is referred to as the 'soluble' protein fraction. The soluble proteins were analyzed immediately by either SDS gelelectrophoresis (Section 2.2.12.2.2) or gelelectrophoresis under non-denaturing conditions (Section 2.2.12.2.1) followed by immunoblotting (Section 2.2.14.1). Analysis of the 'soluble' proteins through by velocity sedimentation through sucrose gradients is described in Section 2.2.21.

##### 2.2.20.2 Preparation of a total leaf protein extract

Either wheat or pea leaf proteins were isolated by grinding 1 g of young tissue (Section 2.2.1.1) with acid-washed sand and 3 ml of isolation buffer (Section 2.2.20.1) plus 2% (w/v) soluble polyvinylpyrrolidone (PVP). The cell-free extract was obtained by centrifugation in an Eppendorf microfuge at 11 600 g for 10 min at 4°C. The supernatant was transferred to a fresh Eppendorf tube kept on ice and was always used the same day.

#### 2.2.21 VELOCITY SEDIMENTATION CENTRIFUGATION

Proteins were analyzed by velocity sedimentation through sucrose gradients by a modification of the method described by Erion *et al.*, (1983). An aliquot (200-500  $\mu$ l) of a desired *E.coli* 'soluble' protein extract (Section 2.2.20.1), to which  $\beta$ -galactosidase had been added (approximately 100 U), was layered onto a 11-ml linear sucrose gradient (7-25% (w/v) in isolation buffer (Section 2.2.20.2), unless indicated otherwise. The gradient was centrifuged in a Beckman L8-70 M ultracentrifuge in a TST41 swinging bucket rotor at 25,000 rpm for 24 h at 4°C, followed by fractionation from the bottom of the gradient. An aliquot (20-50  $\mu$ l) of each fraction (approximately 0,5 ml) was analyzed by SDS polyacrylamide gelelectrophoresis (Section 2.2.12.2.2) followed by immunoblotting (Section 2.2.14). The enzymic activity of the internal marker,  $\beta$ -galactosidase, was directly determined in all the fractions as described in Section 2.2.22.2. Simultaneously, calibration gradients were centrifuged containing purified  $\beta$ -galactosidase (Mr 465.000), aldolase (Mr 160 000) and ovalbumin (Mr 43 000). The calibration gradient fractions were analyzed for Rubisco and  $\beta$ -galactosidase enzyme activity (Section 2.2.22), and the absorbance at 280 nm measured to determine the positions of aldolase and ovalbumin.

#### 2.2.22 ENZYME ASSAYS

##### 2.2.22.1 **Determination of Rubisco activity**

The carboxylase activity of Rubisco was determined either in 'soluble' protein extracts from *E.coli* or total wheat leaf protein extracts as ribulose 1,5-bisphosphate (RuBP)-dependent  $^{14}\text{CO}_2$  fixation, according to a modification of the method described by Lorimer *et al.* (1977). The typical reaction mixture contained 525  $\mu$ l of assay buffer consisting of 100 mM Tris-HCl, 20 mM  $\text{MgCl}_2$ , 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 8.0, 500  $\mu$ l 'soluble' *E.coli* extract (Section 2.2.20.1) and 125  $\mu$ l  $\text{NaH}^{14}\text{CO}_3$  solution consisting of 0.5 ml 40 mM  $\text{NaH}^{14}\text{CO}_3$  (specific activity 1.85 GBq/mmol = 50 mCi/mmol) plus 168 mg

NaHCO<sub>3</sub> dissolved in 9.5 ml distilled water. The reaction mixture was incubated at 30°C for about 2 min and an aliquot (100 µl) of 5 mM RuBP dissolved in assay buffer was added to start the reaction. Incubation at 30°C was continued and the samples (200 µl) taken at desired time intervals were immediately added to 100 µl 12 M formic acid. The samples were dried at 95°C and the residue dissolved in 100 µl distilled water. Scintillation fluid (1 ml), (LKB 'Optiphase' safe) was added and the radioactivity incorporated into the acid-stable product determined by means of a scintillation counter (LKB 1219 Rackbeta).

#### 2.2.22.2 Determination of $\beta$ -galactosidase activity

The method used was originally described by Craven *et al.* (1965) and modified by Miller (1972) and measures the hydrolysis of 0-nitro-phenol- $\beta$ -D-galactose pyranoside (ONPG) to 0-nitro-phenol. An aliquot (10-50 µl) of a sucrose gradient fraction (Section 2.2.21) was added to 1 ml of buffer consisting of 60 mM Na<sub>2</sub>HPO<sub>4</sub>, 5 mM 2-mercaptoethanol, pH 7.0. The reaction mixture was equilibrated at 28°C for 5-10 min and the reaction started by adding 0.2 ml ONPG solution (4 mg/ml in distilled water). After the required incubation time (0-10 min), 0.5 ml of 1 M Na<sub>2</sub>CO<sub>3</sub> was added and the absorbance (A) at 420 nm determined. If desired the enzymic activity was calculated using the following formula:

$$\text{enzymic activity} = \frac{A_{420} \times 380}{T \times V} \quad \text{U/ml}$$

in which: T = time in min of incubation

V = volume of aliquot in ml

### 3. RESULTS AND DISCUSSION

### 3.1 ISOLATION AND CHARACTERISATION OF cDNA SEQUENCES FROM *TRITICUM AESTIVUM*

#### 3.1.1 INTRODUCTION

The isolation of cDNA sequences by means of antibodies was first described by Young & Davies (1983) and requires the construction of a cDNA expression library in a phage vector. A cDNA library representing the mRNA population of green *Triticum aestivum* (wheat) leaf tissue in the phage vector  $\lambda$  gt11 was prepared by C. Raines (Section 2.2.6.1). The expression vector  $\lambda$  gt11 is capable of producing a polypeptide specified by the inserted DNA fragment and can therefore be screened with antibodies to detect DNA sequences encoding the protein against which the antibodies are directed (Young & Davis, 1983a; 1983b).

At the start of this work polyclonal antibodies raised against the Rubisco subunit binding protein purified from *Pisum sativum* (pea) were available (Hemmingsen & Ellis, 1986) and these were used to screen the wheat  $\lambda$  gt11 expression library. The cDNA, isolated from the lysogen SM8 was subcloned in the plasmid vector pGEM3 (Section 3.1.4.1) and further characterized by restriction enzyme analysis (Section 3.1.4.2) and DNA hybridization (Section 3.1.4.3). To determine whether the isolated cDNA could direct the translation of a polypeptide of the correct Mr of about 63,000; i.e. that the cDNA represented a complete coding sequence, transcription and translation analysis *in vitro* was performed (Section 3.1.4.4).

#### 3.1.2 CHARACTERISATION AND ANALYSIS OF THE *PISUM SATIVUM* RUBISCO SUBUNIT BINDING PROTEIN ANTIBODIES.

The antibodies raised against the purified pea Rubisco subunit binding protein (Hemmingsen & Ellis, 1986) were tested for their ability to recognize the Rubisco subunit binding protein from wheat. The purified Rubisco subunit binding protein from pea consists



of two subunits (Hemmingsen & Ellis, 1986) with an apparent Mr of about 61 000 (termed the  $\alpha$  subunit) and 60 000 (termed the  $\beta$  subunit) and are present in equal amounts (Musgrove & Ellis, 1986). Immunologically-related polypeptides to the pea Rubisco subunit binding protein were detected in total protein extracts of wheat leaves. It can be seen in Figure 5, Panel A (lane 2) that the  $\alpha$  and  $\beta$  subunits of the pea Rubisco subunit binding protein are represented by two bands of the same staining intensity. In the wheat leaf extract two bands can be seen, also with the same intensity (Panel A, lane 1), representing polypeptides with an apparent Mr of about 60 000. More precise comparison of the position of these bands with the position of the  $\alpha$  and  $\beta$  subunits of the Rubisco subunit binding protein from pea suggests that the wheat  $\alpha$  and  $\beta$  subunits have a Mr of about 61 500 and 60 500 respectively, i.e. slightly larger than their counterparts in pea. No cross-reacting polypeptides were detected in either the pea or the wheat leaf extract upon incubation with pre-immune serum (Panel B).

In order to use the pea Rubisco subunit binding protein antibodies for the screening of the wheat  $\lambda$  gt11 cDNA expression library, it was essential that all the antibodies which recognized *E.coli* and phage proteins were eliminated. The antibody preparation was therefore purified by incubation with a  $\lambda$  gt11 lysate and analyzed to determine whether the remaining antibodies cross-reacted with the wheat Rubisco subunit binding protein but not with any *E.coli* proteins. Cell extracts of *E.coli* strain TG2, a wheat leaf extract and a partly purified fraction containing the Rubisco subunit binding protein from pea were subjected to SDS polyacrylamide electrophoresis, immunoblotted and the nitrocellulose filters incubated with either the purified or the non-purified Rubisco subunit binding protein antibody preparation from pea. The non-purified antibodies cross-reacted with the  $\alpha$  and  $\beta$  subunits of the Rubisco subunit binding protein from both wheat and pea (Fig. 6, Panel A, lane 2 and 3). In addition a polypeptide in the *E.coli* extract was detectable (Panel A, lane 1) with an apparent Mr of about 60 000. However, incubation with the purified antibodies did not reveal this band in the *E.coli* extract. Sometimes a low Mr band was observed as indicated by the arrow (Panel B, lane 1). This band was a result of the method used to detect the antibody and was also present when the antibody incubation step was omitted from the

Figure 5. Immunological detection of the Rubisco subunit binding protein in extracts of *Triticum aestivum*.

Total protein extracts from pea and wheat leaves were prepared (Section 2.2.20.2) and equal amounts (500  $\mu$ g) of soluble protein were analyzed on a 15% polyacrylamide gel containing SDS followed by immunoblotting. The filter was incubated with, A: serum containing antibodies raised against the purified Rubisco subunit binding protein from pea, B: pre-immune serum. Antibodies were detected using the enzyme-linked method. Lane 1: wheat leaf proteins, lane 2: pea leaf proteins.

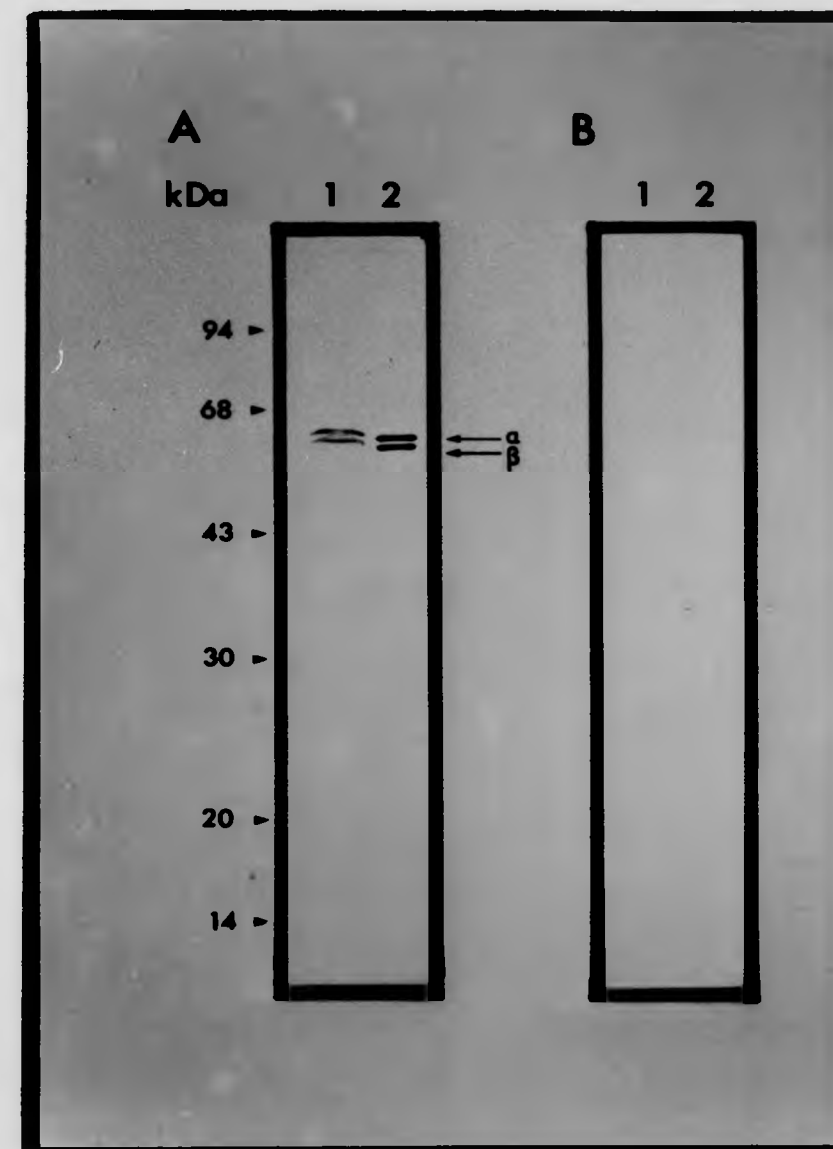
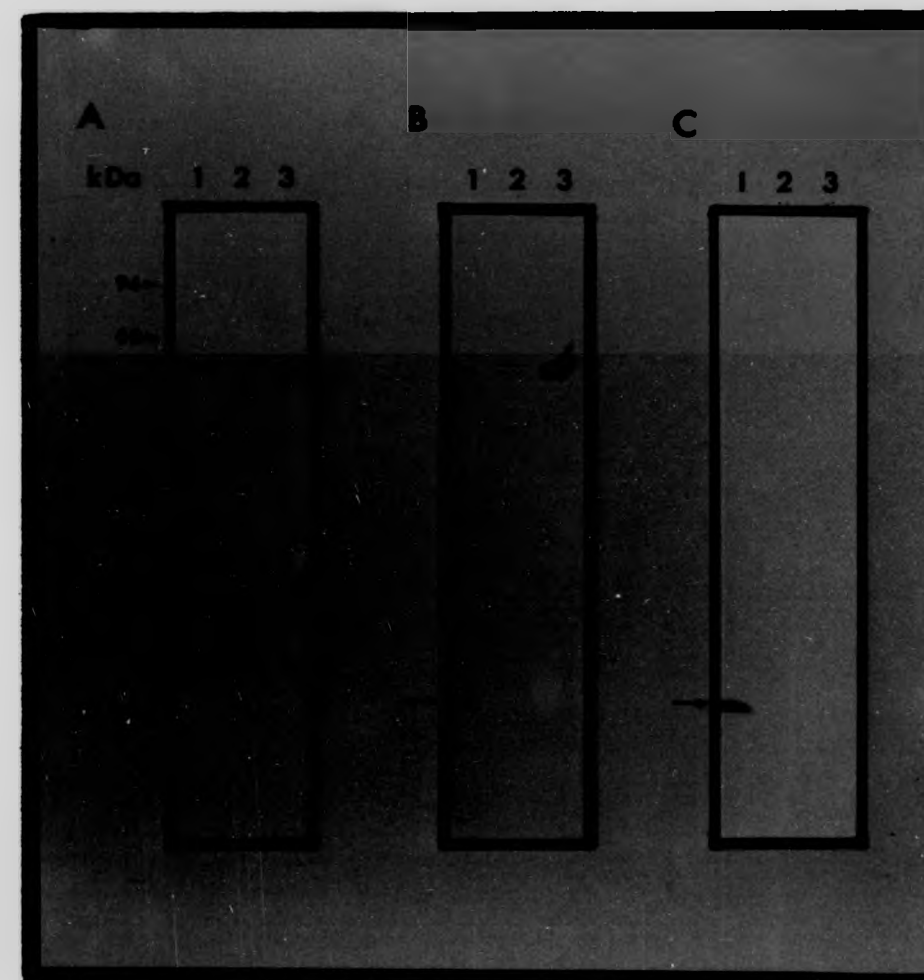


Figure 6. Analysis of the purified Rubisco subunit binding protein antibody preparation.

The soluble protein fractions from wheat leaves and *E.coli* strain TG2 were isolated analyzed on a 15% polyacrylamide gel containing SDS together with a partly purified preparation of the Rubisco subunit binding protein from pea i.e. a fraction of a gelfiltration column enriched in this protein (Hemmingsen & Ellis, 1986). The gel was immunoblotted and the antibodies detected using the coloured-detection method. Panel A: pea Rubisco subunit binding protein antibodies, Panel B: pea Rubisco subunit binding protein antibodies after purification using a  $\lambda$  gt11 lysate, Panel C: no antibodies. Lane 1: *E.coli*, lane 2: wheat, lane 3: pea. A non-specific band is indicated (arrow).



procedure (Panel C, lane 1). As all the antibodies which recognized *E.coli* proteins were eliminated it is very likely that any antibodies which cross-react with phage proteins were also removed. This conclusion was later confirmed when the  $\lambda$ gt11 library was screened (Section 3.1.3).

The purification of the antibody preparation decreased the sensitivity of the detection of the wheat  $\alpha$  and  $\beta$  subunits (Fig. 6: cf. lanes 2 in A and B). The larger decreased detection of the  $\beta$  subunit compared to that of the  $\alpha$  subunit suggests that the  $\beta$  subunit has more antigenic sites in common with the *E.coli* protein than the  $\alpha$  subunit. The cross-reactivity of the pea  $\alpha$  and  $\beta$  subunits was unaffected (Fig. 6 cf. lanes 3 of A and B).

It was concluded that the purified pea Rubisco subunit binding protein antibody preparation was suitable to isolated specific cDNA sequences from the wheat  $\lambda$  gt11 expression library.

### 3.1.3 ISOLATION OF *TRITICUM AESTIVUM* cDNA SEQUENCES FROM THE LAMBDA GT11 EXPRESSION LIBRARY.

The correct expression of any inserted cDNA in the  $\lambda$  gt11 recombinant lysogen will depend on the orientation and reading frame of the inserted DNA with respect to those of the *lacZ* sequence. One-sixth of the  $\lambda$  gt11 recombinants containing a specific cDNA will be predicted to produce a  $\beta$ -galactosidase fusion protein. The proportion of Rubisco subunit binding protein mRNA in the total mRNA pool, as well as the quality of the cDNA, will further affect the ability to detect the antigen produced by the recombinant. A nearly full-length cDNA will produce more potential antigenic determinants than shorter cDNA, and a high titre antibody preparation will obviously produce a better signal than a low titre antibody preparation. According to Huynh *et al.* (1985), it is reasonable to assume that antibodies that produce a good signal on an immunoblot will produce a good signal in the  $\lambda$  gt11 screening procedure.

The  $\lambda$  gt11 expression library, containing  $10^9$  to  $10^{11}$  recombinant phage (Section 2.2.6.1) was screened with the purified Rubisco subunit binding protein antibody

preparation. In a screen of  $6 \times 10^4$  recombinants a total of eight putative lysogens were detected (Fig. 7, Panel B). When the non-recombinant phage was screened with the antibody preparation no signals were observed, confirming that all the antibodies which cross-reacted with any phage  $\lambda$  polypeptides had been removed by the purification procedure (Fig. 7, Panel A). The putative lysogens were subsequently taken through several rounds of plaque purification until all plaques on the plate produced a signal (Fig. 7, Panel C). From the eight putative lysogens two were purified and named SM2 and SM8, while the other six were lost in different rounds of plaque purification.

#### 3.1.4 CHARACTERIZATION OF THE ISOLATED *TRITICUM AESTIVUM* cDNA SEQUENCE

##### 3.1.4.1 Subcloning of the cDNA sequence

The  $\lambda$  gt11 cDNA library had been constructed in such a way that digestion of the isolated recombinant phage DNA with the restriction enzyme *Eco*R1 should release the wheat cDNA sequence (Section 2.2.6.1 and Huynh *et al.*, 1985). To analyze the wheat cDNA, phage DNA was isolated from lysogen SM8 (Section 2.2.5.4). The DNA was digested with *Eco*R1 and analyzed on an agarose gel. It can be seen from Figure 8 that an *Eco*R1 fragment of about 1.9 kb was released. The phage  $\lambda$  vector fragments, 19 and 21 kb each, were not resolved in this gelsystem.

The plasmid vector pGEM3 is a derivative of pBR322 and carries the multiple cloning site arrangement from pUC19 (Yanisch-Perron *et al.*, 1985). In addition, both the SP<sub>6</sub> and T7 RNA polymerase promoters flank the multiple cloning region. This particularly vector was chosen as it provides the possibility of transcription and translation *in vitro* to determine whether the wheat cDNA fragment represents a full-length DNA sequence (see Section 3.1.4.4). The pGEM3 DNA was linearized with *Eco*R1, phosphatased and ligated with the remainder of the SM8 *Eco*R1 digestion mixture. The ligation mixture was transformed into the *E. coli* strain JM109 and ampicillin-resistant white colonies were

Figure 7. Detection of the Rubisco subunit binding protein in  $\lambda$  gt11 lysogens using antibodies.

The  $\lambda$  gt11 non-recombinant and wheat recombinant lysogens were screened with the purified antibody preparation and putative clones were purified. Filters contained  $10^4$  non-recombinant lysogens (A),  $10^4$  wheat  $\lambda$  gt11 lysogens (B) and 25 purified SM8 lysogens (C). The X-ray film was exposed for 5 days with an intensifying screen.

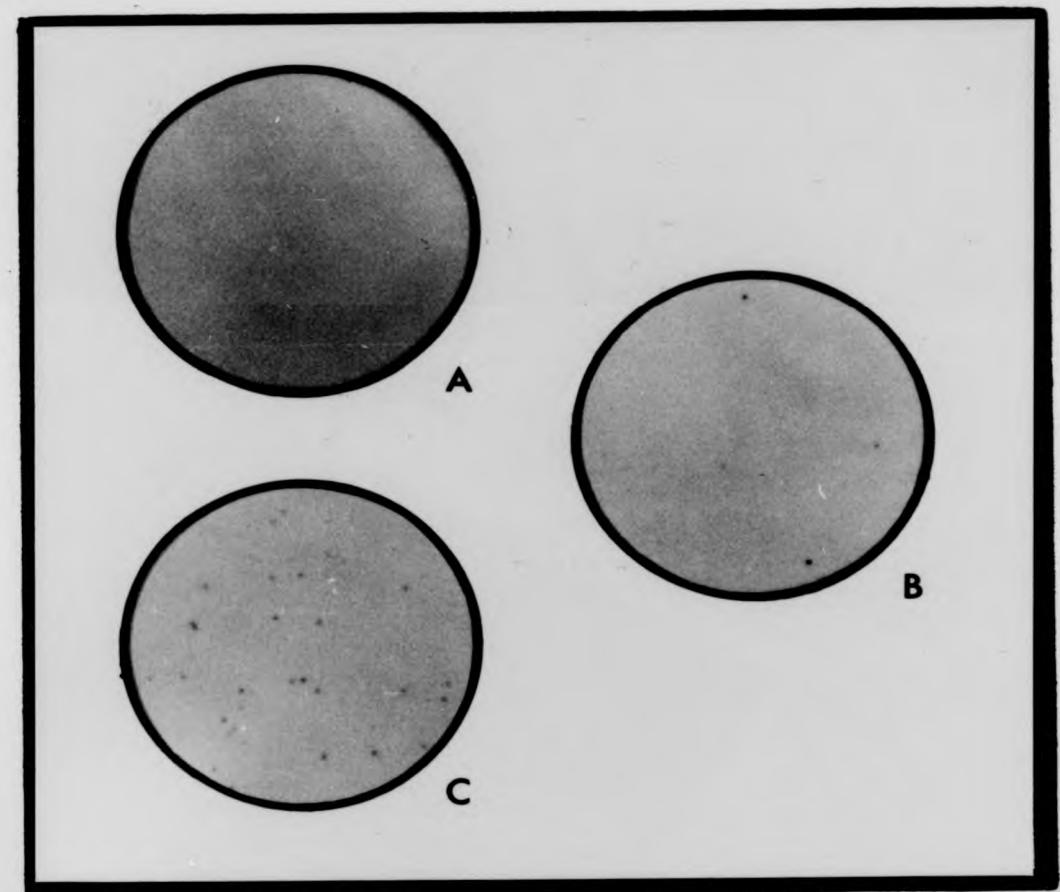
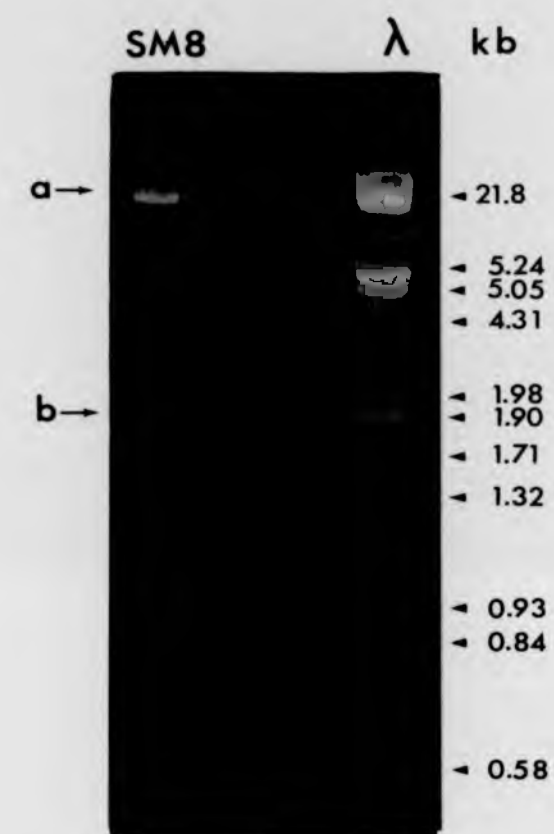


Figure 8. Analysis of the *Triticum aestivum* cDNA from the lysogen SM8.

Phage DNA was isolated from lysogen SM8, digested with *EcoRI* in a total volume of 12  $\mu$ l and an aliquot (5  $\mu$ l) of the mixture analyzed on a 1% (w/v) agarose gel. Phage  $\lambda$  DNA was digested with *EcoRI/HindIII*. The  $\lambda$  gt11 vector fragments of 19 kb and 21 kb (arrow a), and the wheat cDNA fragment of about 2 kb (arrow b), are marked.





selected. Plasmid DNA was isolated from ten selected colonies and digested with *EcoR*I, followed by analysis on an agarose gel to identify those colonies containing the 2 kbp wheat cDNA fragment. Six colonies contained the correct construct and were further analysed by a series of restriction enzyme digestions to distinguish the two possible orientations of the 1.9 kb wheat cDNA fragment.

Digestion of the plasmid DNA with *Sac*I gave two distinct digestion patterns when analyzed on an agarose gel (Fig. 9). The construct in which the *Sac*I site was closest to the SP<sub>6</sub> RNA polymerase promoter was named pSV8, and the construct containing the wheat cDNA fragment in the opposite orientation pSV9 (Fig. 10). The plasmid pSV8 was used for further restriction enzyme digestion analysis.

#### 3.1.4.2 Restriction enzyme digestion analysis of pSV8

Accurate analysis of digestion mixtures of pSV8 with a variety of restriction enzymes was performed on agarose gels which were run for 20 h. The DNA fragments were stained after electrophoresis and the result of a typical analysis is shown in Figure 11 (Panel A). The size of the different DNA fragments was calculated using phage  $\lambda$  DNA digested with *Hind*III and *Eco*RI as markers. A restriction enzyme map was constructed from a series of digestion patterns and is outlined in Figure 11 (Panel B). A more accurate calculation of the wheat cDNA fragment indicated a size of 1.85 kb. No restriction sites were found for the enzymes *Sal*I, *Kpn*I, *Bam*HI, *Pst*I, *Sma*I, *Bgl*II, *Cla*I, *Nde*I and *Xba*I.

#### 3.1.4.3 DNA hybridization analysis of pSV8

At this stage of the work S.M. Hemmingsen (Plant Biotechnology Institute, Saskatoon) reported to us that he had isolated a 1.4 kb cDNA fragment from a *Ricinus communis* (castor bean)  $\lambda$  gt11 cDNA expression library (representing the mRNA population from the endosperm of developing seeds) using the Rubisco subunit binding protein antibodies raised against the protein purified from pea. This DNA fragment contained in plasmid pSBP2 was



Figure 9. Determination of the orientation of the *Triticum aestivum* cDNA in the pGEM3 vector.

The rest of the *EcoRI* digestion mixture (Fig. 8 legend) was used to ligate the wheat cDNA *EcoRI* fragment into the plasmid vector pGEM3, previously digested with *EcoRI* and phosphatased. The ligation mixture was used to transform *E.coli* strain JM109 to ampicillin resistance. Plasmid DNA was isolated from selected white colonies, digested with *EcoRI* or *SacI* and analyzed on a 1% (w/v) agarose gel (Section 2.2.12.1.2.). The lanes contain:  $\lambda$  DNA digested with *EcoRI* and *HindIII*; Panel A pGEM3 vector DNA (lane 1), pSV8 (lane 2) and pSV9 (lane 3) all digested with *EcoRI*; Panel B, pGEM3 vector DNA (lane 1), pSV8 (lane 2) and pSV9 (lane 3) all digested with *SacI* (see also Fig. 10).

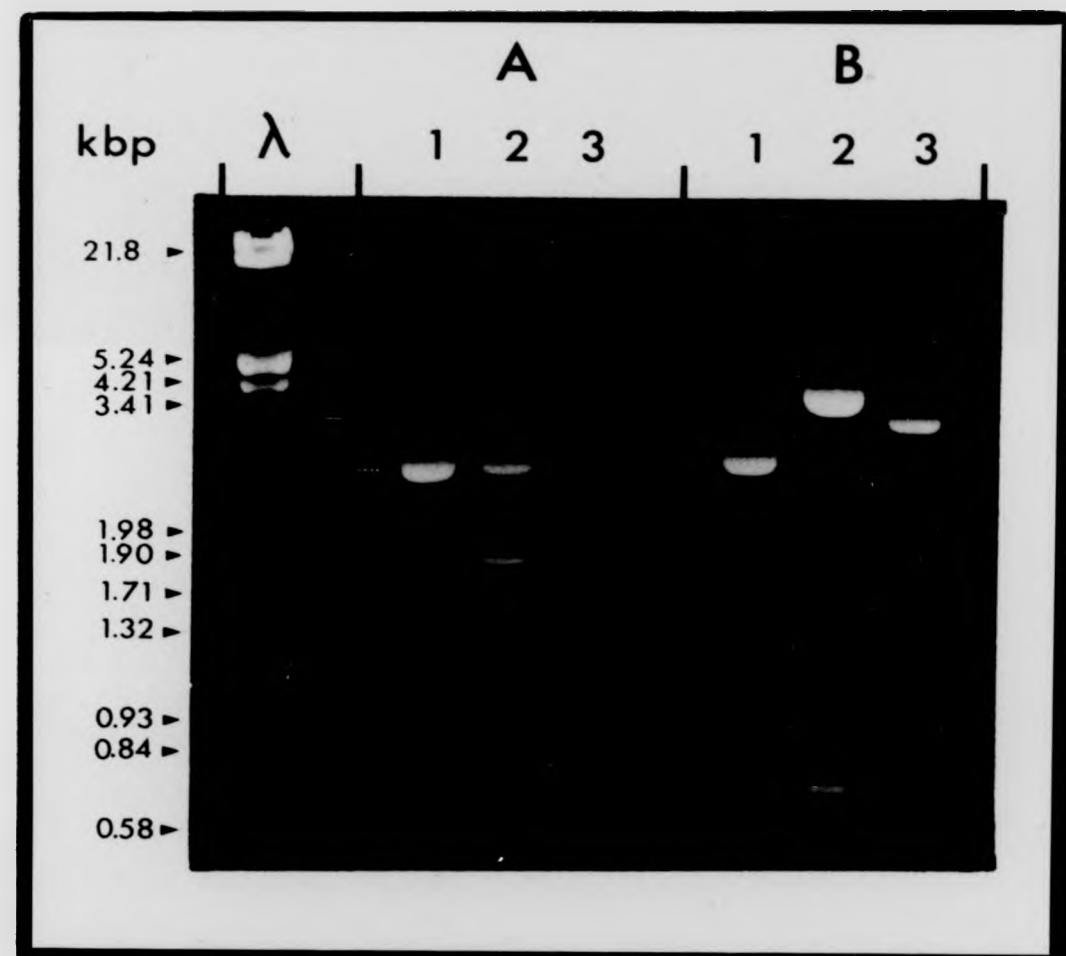


Figure 10. Structure of plasmids pSV8 and pSV9.

The plasmid pSV8 contains the wheat cDNA sequence in which the *SacI* site is situated closest to the phage SP<sub>6</sub> promoter. The construct in which the wheat cDNA sequence is present in the opposite orientation is called pSV9. The restriction enzyme sites for *EcoRI* (E) and *SacI* (S), the multiple cloning site (mcs) and the gene for ampicillin resistance (Amp) are indicated. Transcription from the different promoters occurs in the directions shown (arrows).

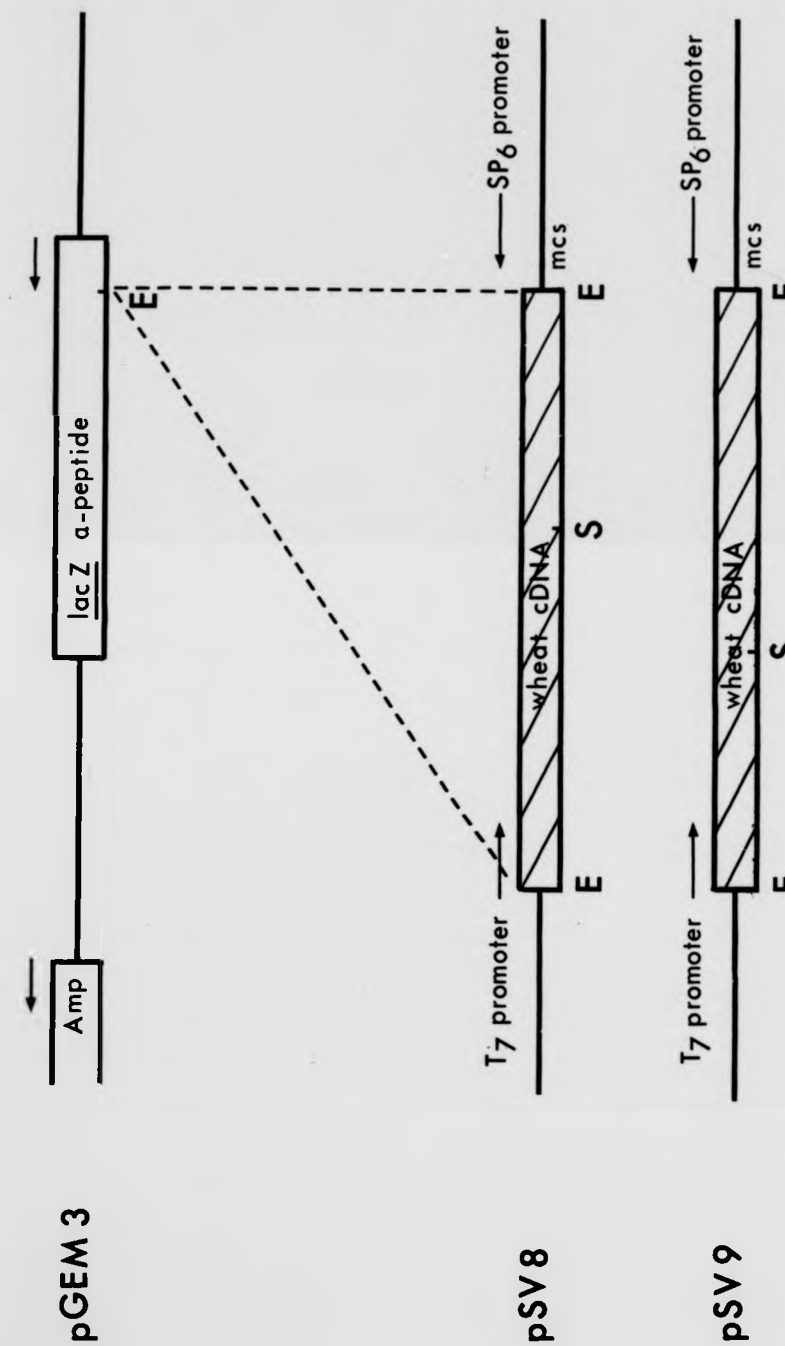
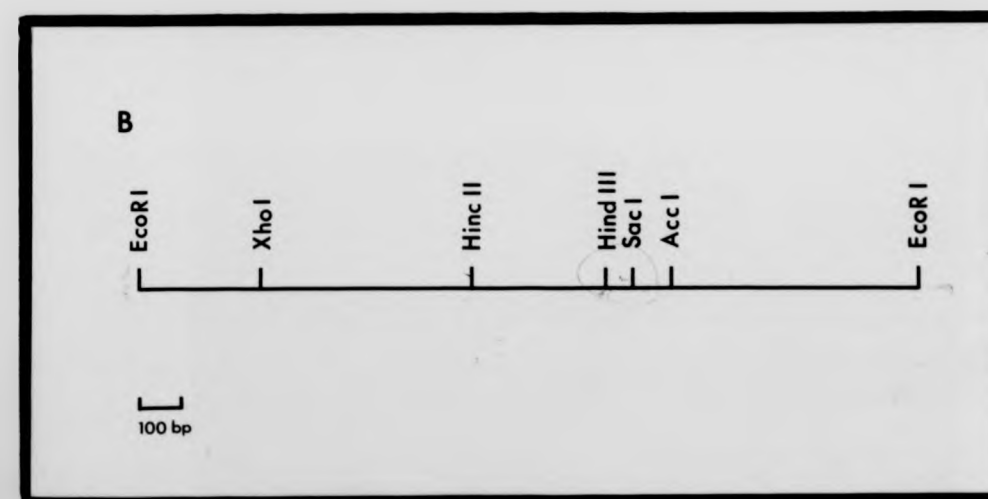
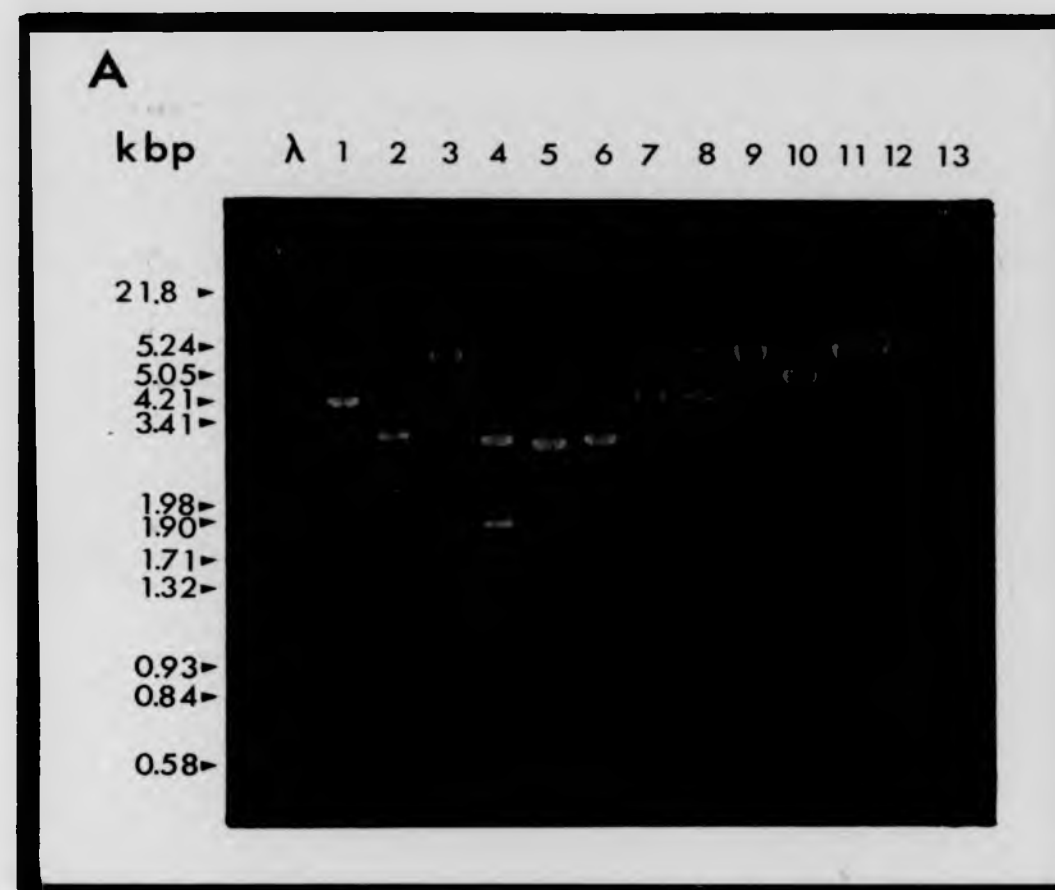


Figure 11 Determination of restriction enzyme sites in the *Triticum aestivum* cDNA sequence.

Vector pGEM3 DNA and plasmid pSV8 DNA were digested with a variety of restriction enzymes (Section 2.2.7.1) and the generated DNA fragments were separated electrophoretically in a 1% (w/v) agarose gel (Section 2.2.12.1.1.). Panel A: Lane ( $\lambda$ ),  $\lambda$  DNA digested with *EcoRI/HindIII*; lane (1), pGEM3 undigested; lane (2), pGEM3 digested with *EcoRI*; lane (3), pSV8 undigested; lane (4), pSV8 digested with *EcoRI*; lane (5), *EcoRI/HindIII*; lane (6), *EcoRI/SacI*; lane (7), *HindIII*; lane (8), *KpnI*; lane (9), *PstI*; lane (10), *SacI*; lane (11), *SalI*, lane (12), *SmaI* and lane (13), *XbaI*.

Panel B shows a constructed restriction enzyme map of the wheat cDNA sequence (total length of 1.85 kb).



generously given to me and was used for hybridisation with the isolated wheat cDNA fragment.

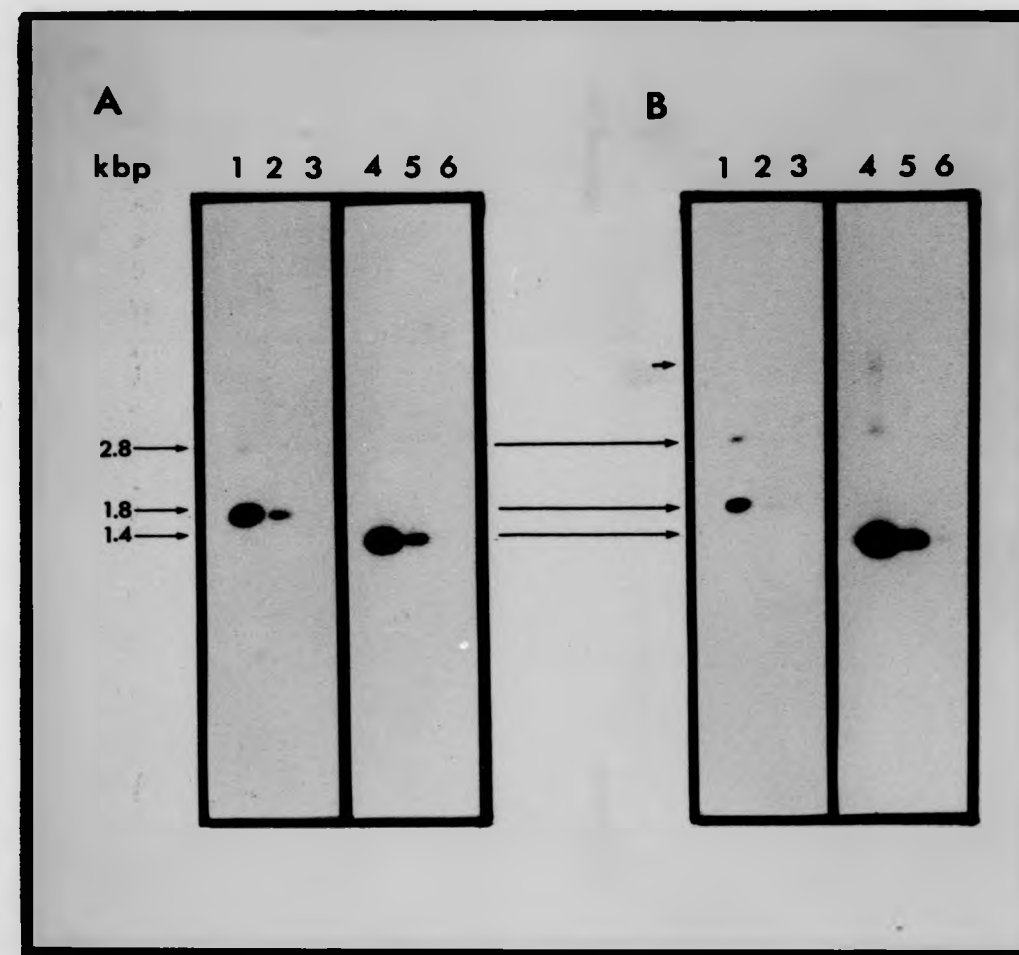
The plasmid pSV8 was digested with *EcoRI* and the separated DNA fragments transferred to nitrocellulose. The radiolabelled purified 1.4 kb castor bean cDNA fragment was used for hybridisation. The results of two washes, one at low and one at high stringency are shown in Figure 12. After a low stringency wash (Panel A), corresponding to a 34% base pair mismatch between the castor bean and the wheat DNA sequences (Table 1), a strong hybridisation signal was observed. As equal amounts of wheat and castor bean DNA were loaded on the gel it was concluded that the homologous hybridisation was stronger than the heterologous hybridisation (Panel A, cf. lanes 4, 5 and 6, and lanes 1, 2 and 3). After washing the filter under high stringency conditions, corresponding to about 1% bp mismatch between the two sequences (Table 1), a strong hybridisation signal was still visible (Panel B), suggesting that the wheat and the castor bean sequences are very similar. The stronger signal observed for the homologous hybridisation after the high stringency wash (Panel B, lanes 4, 5 and 6) compared to that after the low stringency wash (Panel A, lanes 4, 5 and 6) is a reflection of the longer exposure time of 20 h (Panel B) compared to that of 5 h (Panel A). After the high stringency wash, the homologous hybridisation (Panel B, lanes 4, 5 and 6) was about 10-fold stronger than the heterologous hybridisation (Panel B, lanes 1, 2 and 3). The faint bands that are present on the autoradiograph at positions corresponding to about 2.8 and 4.5 kb, represent linearized vector DNA and undigested plasmid respectively, implicating that the hybridising DNA probe contained some vector DNA.

#### 3.1.4.4 Transcription and translation of the *Triticum aestivum* cDNA sequence *in vitro*

The Rubisco subunit binding protein  $\alpha$  and  $\beta$  subunits are synthesized as precursors from their nuclear-encoded genes on cytoplasmic ribosomes (Section 1.3.3.; Hemmingsen & Ellis, 1986). The apparent  $M_r$  of the mature  $\alpha$  and  $\beta$  subunits is about 60 000, and the

Figure 12. DNA hybridisation of the *Triticum aestivum* and the *Ricinus communis* cDNA sequences.

Plasmid DNA of pSV8 and pSBP2 was digested with *Eco*RI (Section 2.2.7.1) and the mixtures analyzed by electrophoresis through a 1% (w/v) agarose gel (Section 2.2.12.1.1.). pSBP2 contained a 1.4 kbp cDNA fragment in the plasmid vector pGEM3, and was isolated by S.M. Hemmingsen from a *Ricinus communis* (castor bean) endosperm expression library in  $\lambda$  gt11 using the antibodies raised against the purified Rubisco subunit binding protein from pea. The DNA fragment were transferred to nitrocellulose and the filter prehybridised at 55°C for 2 h, followed by hybridisation with the radiolabelled isolated 1.4 kbp castor bean DNA fragment at 55°C for 30 h. The filter was washed for 2 h at 60°C in 5 x SSC containing 0.1% (w/v) SDS (2 x 250 ml), covered with Cling film and exposed to X-ray film for 5 h with an intensifying screen (Panel A). The filter was then washed for 2 h at 65°C in 0.1 x SSC containing 0.1% (w/v) SDS (2 x 250 ml), dried and exposed to X-ray film for 30 h with an intensifying screen (Panel B) lane 1, pSV8, 200 ng; lane 2, pSV8, 20 ng; lane 3, pSV8, 2 ng; lane 4, pSBP2, 200 ng; lane 5, pSBP2, 20 ng; lane 6, pSBP2, 2 ng. The 2.8 kbp fragment represents linearized vector DNA and undigested plasmid DNA is indicated by the short arrow.



precursor form for both the subunits from pea has an apparent Mr of about 63 000 (Hemmingsen & Ellis, 1986). A cDNA fragment encoding a precursor polypeptide would thus have an estimated length of about 2090 to 2190 bp, including a 200 to 300 bp untranslated region (average Mr per amino acid used was 100). Translation *in vitro* of the wheat cDNA sequence to give a polypeptide of the correct Mr, i.e. Mr of about 63 000 would suggest that the isolated wheat DNA sequence encodes the precursor for the  $\alpha$  or  $\beta$  subunit polypeptide.

Both constructs, pSV8 and pSV9, were used to produce RNA *in vitro*, as it was unknown which plasmid contained the wheat cDNA sequence in the correct orientation for translation. The highly purified template was linearized to completion to avoid the synthesis of transcripts containing plasmid sequences (Krieg & Melton, 1987). The restriction enzyme *SalI* was chosen since this restriction site is not present in the wheat cDNA sequence and the enzyme cuts at a site in the polylinker region, leaving the T7 RNA polymerase transcription initiation site right next to the wheat sequence. This site is referred to as the 5' end of the transcript, whereas the 3' end of the transcript is determined by the site at which the DNA is cleaved with the restriction enzyme. The T7 RNA polymerase was used for transcription of the wheat sequences and radioactively labelled transcripts were analysed on agarose gels containing formaldehyde, followed by autoradiography. Both the constructs, pSV8 and pSV9, produced one transcript of the expected size of about 1.85 kb (Fig. 13).

Unlabelled transcripts from pSV8 and pSV9 were used for translation *in vivo*. The translation products, labelled with [ $^{35}\text{S}$ ]methionine, were analyzed on SDS polyacrylamide gels followed by autoradiography. Two polypeptides, with an apparent Mr of about 55 000 (P55) and 59 000 (P59), were observed using the pSV8 transcript (Fig. 14, lane 2), but no polypeptides in this range were detectable with the pSV9 transcript (Fig. 14, lane 3). The polypeptide of Mr of 59 000 (marked P1 in Fig. 14) is too small to represent a precursor for either the  $\alpha$  or the  $\beta$  subunit. The smaller polypeptide of Mr 55 000 (marked P2 in Fig. 14) is either a degradation product of Mr 59 000 as a result of proteolytic activity in the wheat germ translation extract or a result of the initiation of translation at internal sites of the mRNA (see also Section 3.2.3.1 and Section 3.2.6.1).



Figure 13. Transcription *in vitro* of the *Triticum aestivum* cDNA sequence.

Plasmid DNA of pSV8, pSV9 and pSBP2 was digested to completion with *Sa*II and used as a template to synthesize radiolabelled RNA *in vitro* (Section 2.2.17.1). The RNA was analyzed on a 1% (w/v) agarose gel containing formaldehyde, the gel dried and exposed to X-ray film for 30 h with an intensifying screen. Lane 1, pSV8; lane 2, pSV9 and lane 3, pSBP2. The *Triticum aestivum* RNA (arrow a) and the *Ricinus communis* RNA (arrow b) are indicated.

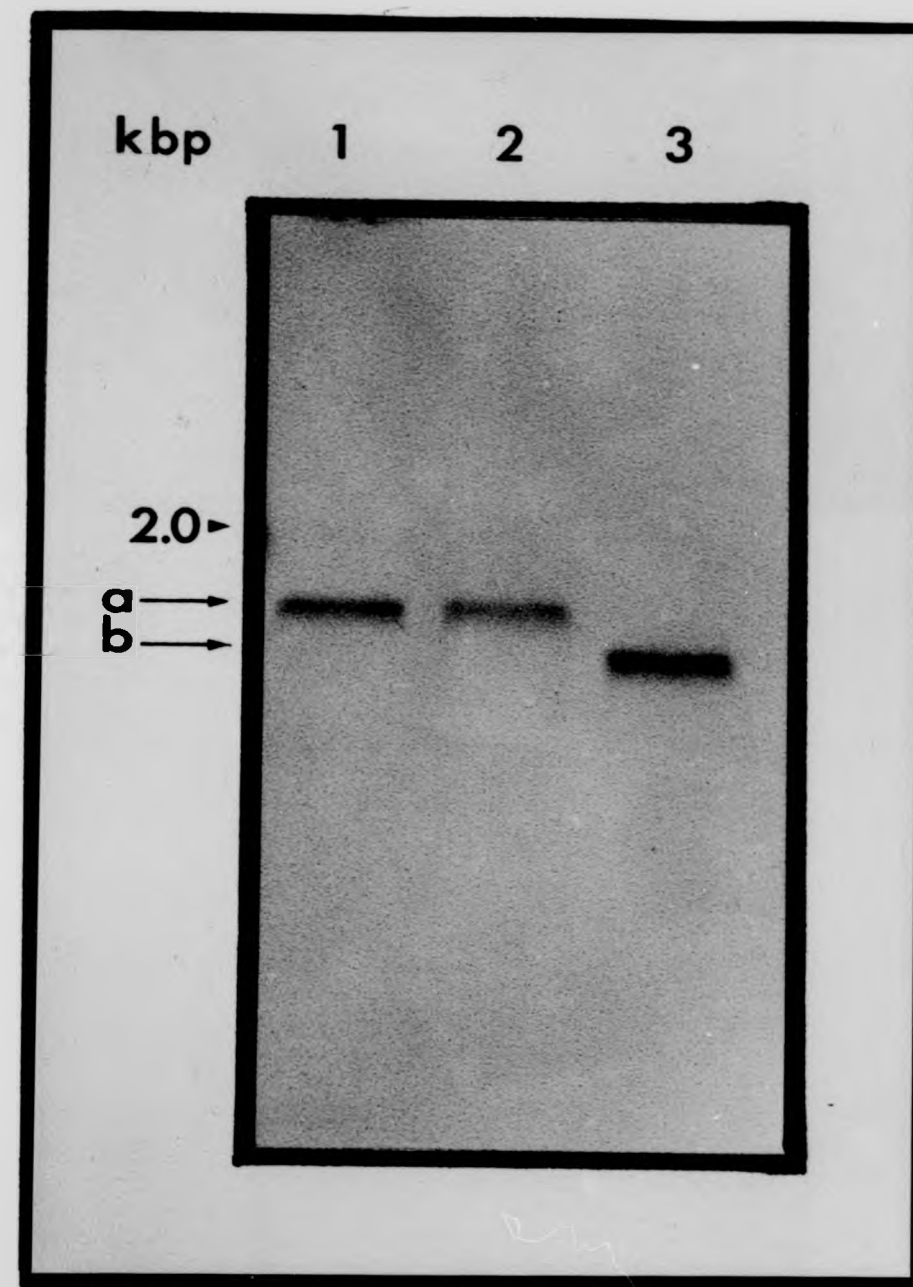
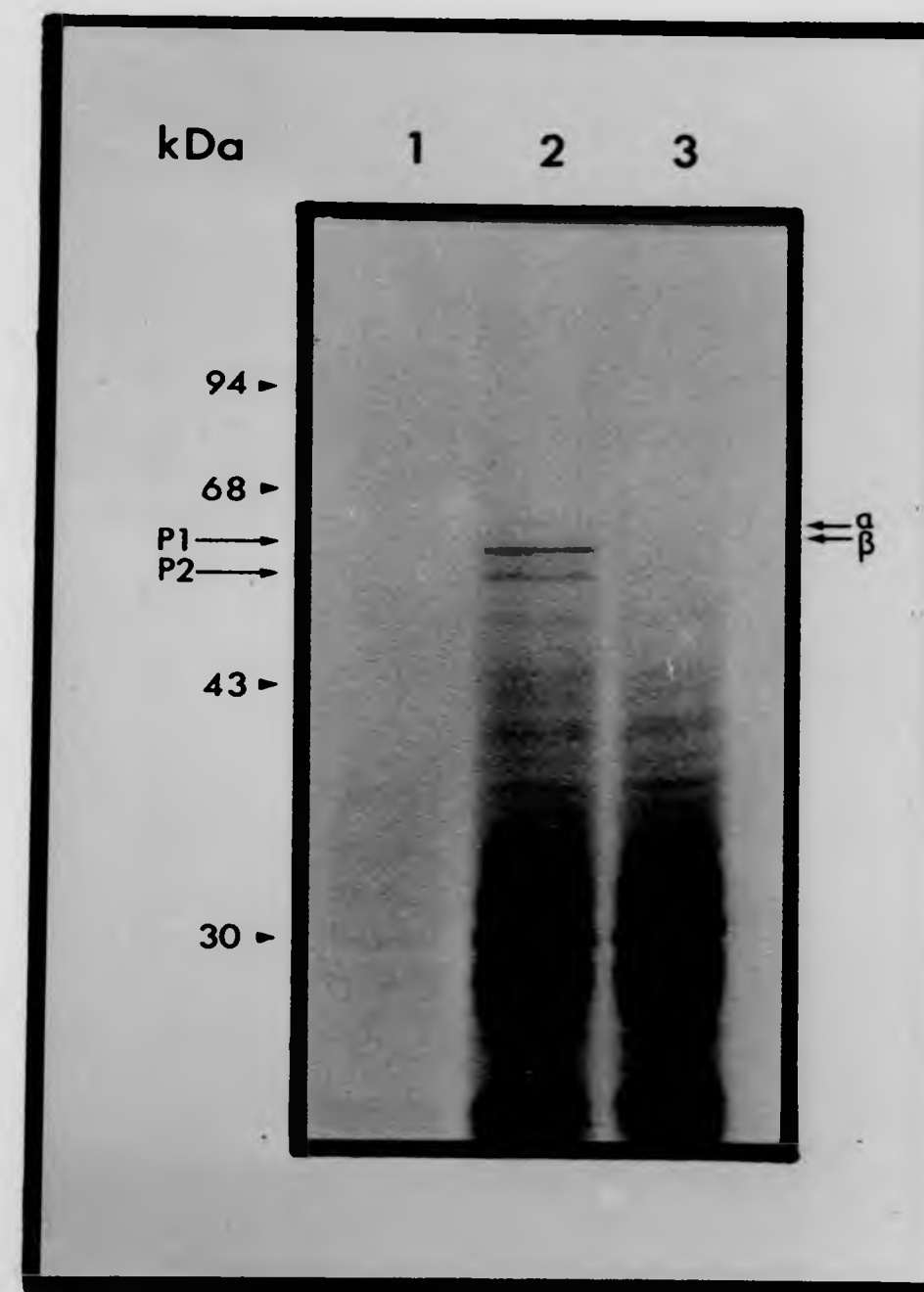


Figure 14. Analysis of translation products from pSV8 and pSV9 transcripts.

Unlabelled transcripts of pSV8 and pSV9 were translated in a wheat germ extract in the presence of [ $^{35}\text{S}$ ] methionine (Section 2.2.17.2), and the extracts analyzed on a 15% (w/v) polyacrylamide gel containing SDS. The amplified and dried gel was exposed to X-ray film for 2 days with an intensifying screen. Lane 1: no transcript, lane 2: pSV8 transcript, lane 3: pSV9 transcript. The translation products in lane 2 had apparent Mr of about 59.000 (P1) and 55.000 (P2). The positions of the purified Rubisco subunits from *P.sativum* are indicated ( $\alpha$  and  $\beta$ ).





### 3.1.5 DISCUSSION

Green wheat leaves contain two abundant soluble polypeptides that cross-react immunologically with the antibodies raised against the purified Rubisco subunit binding protein from pea. These polypeptides have apparent  $M_r$  values of about 61 500 and 60 000 and were named  $\alpha$  and  $\beta$  respectively following the terminology used for the pea Rubisco subunit binding protein subunits (Musgrove & Ellis, 1986). The purified Rubisco subunit binding protein from pea and barley is an oligomeric protein consisting of equal amounts of the two subunits (Musgrove & Ellis, 1986; Johnson, 1987). The Rubisco subunit binding protein antibodies cross-react equally with the  $\alpha$  and  $\beta$  subunits in a pea leaf extract (Fig. 5) and in the purified pea oligomeric protein (Musgrove *et al.*, 1987). The immunological cross-reactivity of the two subunits in a wheat leaf extract is equal to each other, but lower than that of the  $\alpha$  and  $\beta$  subunits in a pea leaf extract (Fig. 5), indicating that either, not all the antibodies recognize the wheat Rubisco subunit binding protein, or the amount of Rubisco subunit binding protein is less per amount of total soluble leaf protein in wheat than in pea, as equal amounts of soluble protein were loaded on the gel. The possibility that either one or both of the wheat subunits is present in lower amounts, but cross-reacts more strongly with the antibodies than the pea subunits is unlikely, but cannot be excluded. Purification of the pea Rubisco subunit binding protein antibodies to eliminate those antibodies that cross-react with *E.coli* and  $\lambda$  phage proteins reduces the immunological reactivity with the  $\alpha$  and  $\beta$  subunits in wheat leaf extracts by a considerable factor, whereas the immunological reactivity with the pea subunits is not affected (Fig. 6). The above observations suggest that in wheat, the amount of Rubisco subunit binding protein per total amount of soluble leaf protein is of the same order as that in pea and that the  $\alpha$  and  $\beta$  subunits are probably present in equal quantities, but do not show such a strong cross-reactivity with the antibodies as the subunits from pea.

Because the immunological signal for the wheat  $\beta$  subunit is clearly less than that for the  $\alpha$  subunit (Fig. 6, Panel B, lane 2), the chance of isolating an  $\alpha$  subunit from the  $\lambda$  gt11

expression library is greater. In a screen of  $6 \times 10^4$  recombinants eight putative lysogens were detected with the antibody and two were purified to homogeneity. The low titre of the antibody preparation resulted in a long exposure of the X-ray film (5 days or more) and consequently the long storage of the plates containing the lysogens (up to 10 days) increased the chance of contamination. The phage DNA isolated from SM2 contained a wheat cDNA insert of about 400 bp (data not shown) and was not further analyzed.

The isolated wheat cDNA sequence from SM8 is about 1.85 kb in size and contains a number of restriction enzyme sites as shown in Figure 11, Panel B. This cDNA fragment shows a high degree of similarity to a 1.4 kbp fragment that was isolated by S.M. Hemmingsen from a castor bean  $\lambda$  gt11 expression library with the aid of the same Rubisco subunit binding protein antibodies. Strong hybridisation is observed under stringent conditions, corresponding to a mismatch of about 1% between the two sequences, suggesting a DNA sequence similarity > 90%. The determined amino-terminal sequences of the purified mature  $\alpha$  and  $\beta$  polypeptides of the Rubisco subunit binding protein from pea, share 11 identical amino acids out of 28, but only when extra residues are introduced in the sequences to optimize the alignment (Ellis *et al.*, 1987). This suggests that the similarity of the  $\alpha$  and  $\beta$  cDNA sequences is probably about 40-50%. The observed similarity between the wheat and the castor bean DNA sequences of >90% indicates that these sequences represent the same Rubisco subunit binding protein subunit i.e. either the  $\alpha$  or the  $\beta$  polypeptide.

Translation of the single wheat cDNA transcript *in vitro* produced two polypeptides with apparent Mr values of about 59 000 (P59) and 55 000 (P55). The expected minimum Mr of an  $\alpha$  or  $\beta$  precursor is about 63 000 (Section 3.1.4.4), indicating that the 1.85 kbp wheat cDNA fragment does not represent a full length cDNA sequence. The P55 polypeptide is either a proteolytic product or the result of internal initiation of translation or the translation product of a short incomplete transcript. The latter is unlikely however since only one RNA species was observed (Fig.11). A small undetectable quantity of truncated RNA can account for the amount of P55 synthesized only if it is assumed that the truncated RNA is translationally more active than the longer RNA species, since equal amounts of P55 and

P59 protein were detected, and this possibility seems unlikely. The synthesis of smaller polypeptides from a given transcript was also observed by Chang *et al.* (1988), who reported that translation of the cDNA coding for the human androgen receptor in a rabbit reticulocyte lysate resulted in synthesis of seven polypeptides of various lengths, which are all immunoprecipitable with a human serum containing autoimmune antibodies to the androgen receptor. Analysis of this DNA sequence showed that internal ATG codons were used to produce four of the smaller polypeptides. Translation of transcripts from *Xenopus* cDNA in a wheat germ extract resulted in the synthesis of more than one polypeptide, all smaller than the expected maximum size. (R.W. Old, personal communication).

Subsequent analysis of the determined wheat cDNA sequence shows that P55 cannot be a result of internal initiation of translation; this point is discussed in detail in Section 3.2.6.1. It is concluded that P55 is probably the result of proteolytic activity in the wheat germ extract.

## 3.2 NUCLEOTIDE SEQUENCE DETERMINATION AND ANALYSIS OF THE ISOLATED *TRITIMUM AESTIVUM* cDNA

### 3.2.1 INTRODUCTION

To determine the nucleotide sequence of the isolated wheat cDNA, double-stranded DNA was used in a modification of the dideoxy chain termination method of Sanger *et al.* (1977), as described by Murphy & Kavanagh (1988). The use of specially designed vectors such as pUBS or Bluescript, which contain a number of different primer sites and a polylinker region with unique restriction sites to aid unidirectional deletion by Exonuclease III, makes nucleotide sequence determination simple and rapid (Section 3.2.2 and Fig. 16). The determined DNA sequence was analyzed for characteristic sequences involved in transcriptional termination and polyadenylation (Section 3.2.3). Translation of the nucleotide sequence revealed the presence of a single open reading frame (Section 3.2.3) and the predicted aminoterminal, amino acid sequence was compared with known sequences of the Rubisco subunit binding protein  $\alpha$  and  $\beta$  subunits from *Pisum sativum* and *Triticum aestivum* (Section 3.2.4). The amino acid sequence analysis included codon usage, determination of the hydrophobicity/hydrophilicity profile, secondary structure prediction and a search for specific binding sites possibly present in the sequence (Section 3.2.5). The results are discussed in terms of the known properties of the polypeptide (Section 3.2.6).

### 3.2.2 DETERMINATION OF THE *TRITICUM AESTIVUM* cDNA SEQUENCE

#### 3.2.2.1 Construction of pSV10 and pSV11

The wheat cDNA sequence was transferred to the plasmid vector pUBS, which differs from the Bluescript vector in that the M13 origin of the replication has been deleted; this deletion abolishes the use of the *Dra*II site in the polylinker region

(G. Murphy, personal communication). This polylinker region is designed to allow specific restriction enzyme digestion which generates DNA ends that are either susceptible or resistant to treatment with exonuclease III. In addition pUBS contains a total of six different primer regions, three on each site of the polylinker, which makes it possible to determine the nucleotide sequence of the inserted DNA starting from opposite ends.

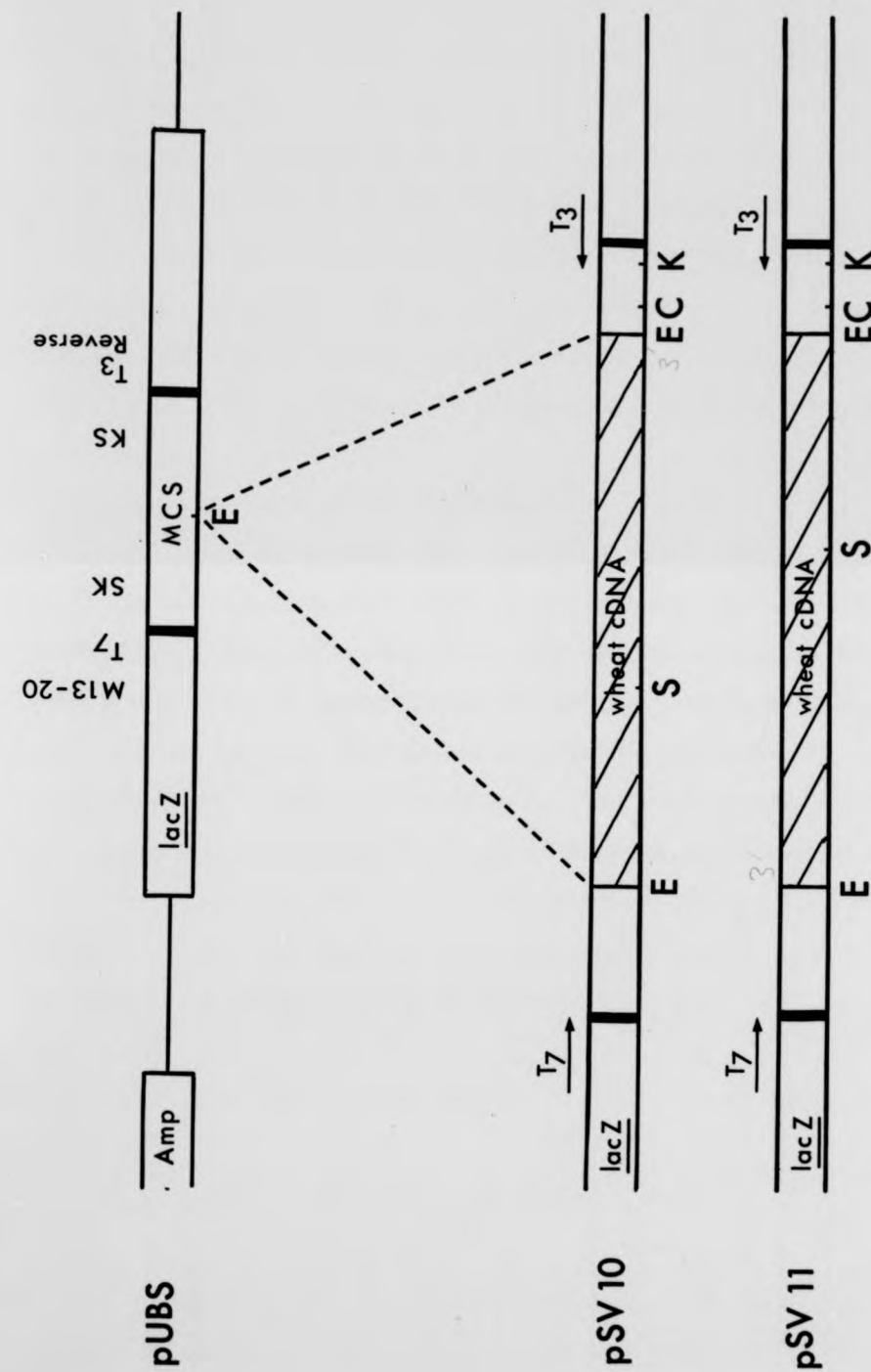
The wheat cDNA was isolated as an *EcoRI* fragment from pSV8 (Fig. 10) and cloned into the *EcoRI* site of the plasmid pUBS, to yield pSV10, in which the asymmetrical *SacI* site is closest to the T7 primer region. The construct containing the *EcoRI* fragment in the opposite orientation was named pSV11 (Fig. 15).

#### 3.2.2.2 Generation of *Triticum aestivum* cDNA deletion derivatives

The production of unidirectional deletions of double-stranded DNA with the aid of exonuclease III is well-suited for the sequencing of long stretches of DNA and was first described by Henikoff (1984). The method involves the progressive digestion of double-stranded DNA at a very uniform rate by exonuclease III and takes the advantage of the enzymes requirement i.e. it will not digest 3' single-strand overhangs, but will digest 3' ends from blunt or 5' overhangs. The polylinker region in pUBS contains unique restriction sites which after digestion leave a 3' single overhang, while other restriction sites leave a blunt or 5' overhang. To create unidirectional deletions in the wheat cDNA sequence pSV10 and pSV11 were digested first with *KpnI*, leaving a 3' overhang adjacent to the sequencing primer, and subsequently digested with *ClaI*, generating a 5' overhang adjacent to the wheat cDNA insert. The restricted DNA was treated with exonuclease III and aliquots removed at timed intervals. Subsequent S1 nuclease digestion resulted in a series of deleted insert derivatives with one common end adjacent to the sequencing primer, and was followed by the addition of DNA polymerase I to ensure the formation of blunt ends. Analysis of the wheat cDNA taken at different time intervals showed the removal of about 180 bp/min by exonuclease III.

Figure 15. Structure of plasmids pSV10 and pSV11

The plasmid pSV10 was generated by transferring the wheat cDNA fragment to the *EcoRI* site present in the pUBS vector such that the asymmetrical *SacI* site is closest to the T7 phage promoter. The construct containing the cDNA in the opposite orientation is named pSV11. The restriction enzyme sites *EcoRI* (E), *SacI* (S), *ClaI* (C) and *KpnI* (K), the multiple cloning site (MCS), the gene for ampicillin resistance (Amp) and the different primer sites (M13-20, T7, SK, KS, T3 and reverse) are indicated. The direction of transcription is shown by the arrows.



DNA ligase was used to circularize the derivatives which were used for transformation and plating (Fig. 16 and Fig. 17). DNA isolated from transformants of sequential plates was digested with *Pvu*II and analyzed by gel electrophoresis to identify the wheat cDNA deletions. Not all the deletions generated were those expected purely on the basis of the exonuclease III digestion time (Fig. 18). DNA analysis of ten transformants selected from one particular time point showed that about one-third of the deletions had an unexpected size. It is possible that the observed degradation was caused by pre-existing nicks in the plasmid DNA. Out of a total of 250 transformants analyzed, from both pSV10 and pSV11, 193 (i.e. 77%) could be used for nucleotide sequence determination.

#### 3.2.2.3 Nucleotide sequence determination

The nucleotide sequence was determined using the method described by Murphy & Kavanagh (1988). To determine whether the wheat cDNA fragment contained a polyA tail, pSV10 was utilized for sequencing using the T3 and T7 primers present on either site of the wheat cDNA sequence. A short poly(A) region of eight As was found closest to the T7 primer site and can be seen in Figure 19, which shows a typical sequencing gel. The complete nucleotide sequence of both the DNA strands was obtained using a selected range of pSV10 and pSV11 deletion derivatives and the T3 primer as a template for DNA polymerase. The determined nucleotide sequence of the wheat leaf cDNA fragment and the derived amino acid sequence are presented in Figure 20.

### 3.2.3 ANALYSIS OF THE DNA SEQUENCE

#### 3.2.3.1 Identification of the translation initiation site

The translation of the wheat cDNA transcript *in vitro* revealed that two polypeptides are synthesized with apparent *M<sub>r</sub>*s of about 59 000 and 55 000 as judged by SDS polyacrylamide gelelectrophoresis (Section 3.1.4.4); the synthesis of two



Figure 16. Outline of the strategy for targeting deletions.

For nucleotide sequence determination aliquots were removed from the exonuclease III reaction mixture and separately processed at timed intervals of 40 s until almost the entire insert was digested (Henikoff, 1984). The ligated DNA was used to transform MC1022 to ampicillin resistance. Symbols are as in Fig. 15.

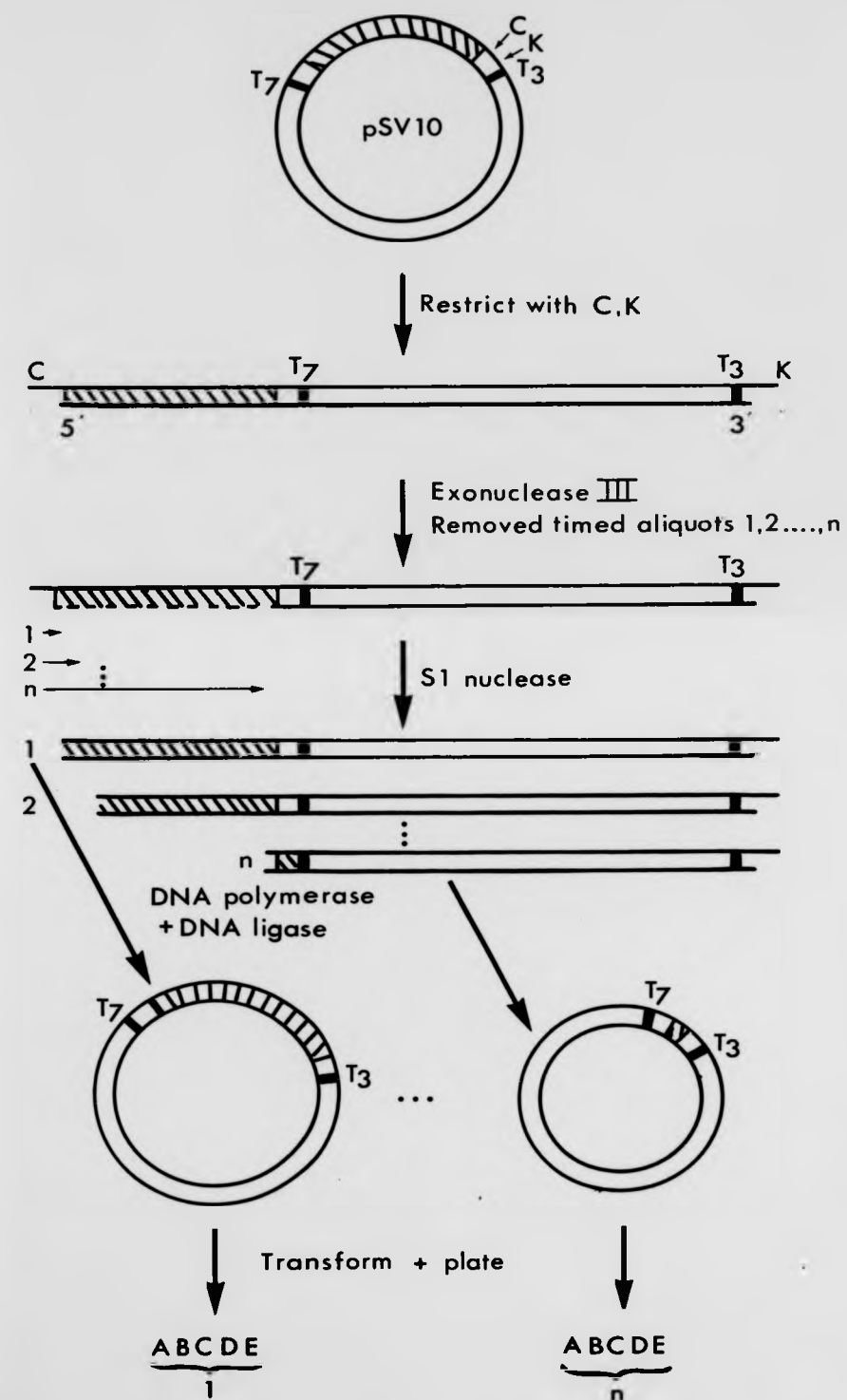




Figure 17. Timecourse analysis of pSV11 deletions.

An aliquot (5  $\mu$ l) of the Exonuclease III reaction mixture was added to 5  $\mu$ l TE<sub>10</sub> (Section 2.2.7.5.). Half of the resulting mixture (5  $\mu$ l) was digested with *Eco*RI and the digestion mixture analyzed on a 1% agarose gel. The numbers indicated different time points: 0 = 0 s, 1 = 80 s, 2 = 160 s, 3 = 240 s, 4 = 320 s and 5 = 400 s,  $\lambda$ : lambda DNA digested with *Ava*II. The *Eco*RI cDNA fragment (1.85 kb) and the pUBS vector (V) are indicated by arrows.

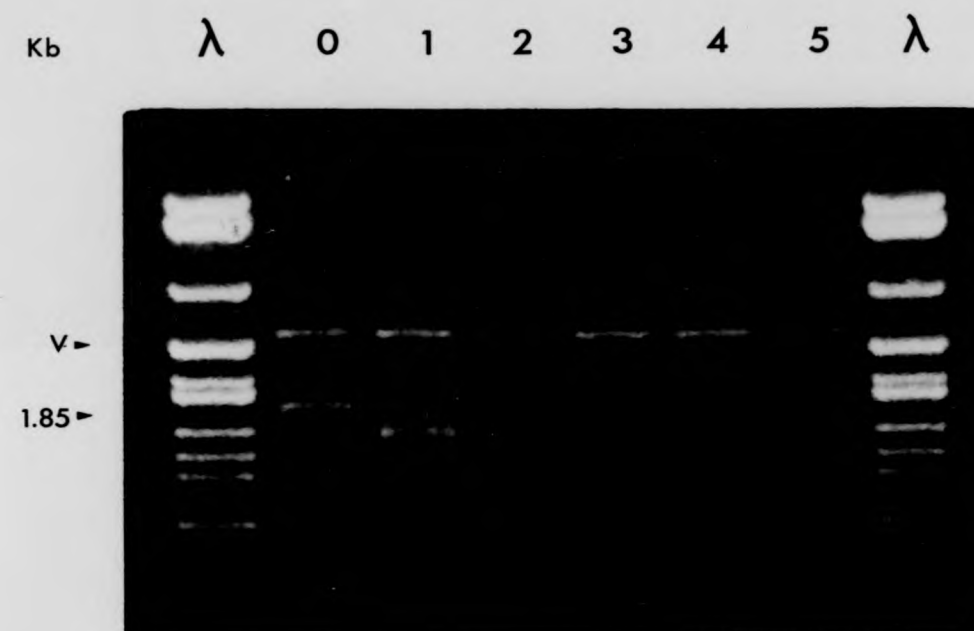


Figure 18. Analysis of pSV11 deletion transformants

Transformants (three for every time point) were grown in LB medium containing ampicillin (50  $\mu$ l/ml) for 30 h and used as a source of plasmid DNA (Section 2.2.5.3.2.). The final DNA pellet was dissolved in 30  $\mu$ l of T10 E0.1 and an aliquot (1  $\mu$ l) digested with *Pvu*II. Separation of the DNA fragments in a 1% (w/v) agarose gel is shown. The numbers 1 to 7 correspond to the different time points; 1 = 40 s, 2 = 80 s, 3 = 120 s, 4 = 160 s, 5 = 200 s, 6 = 240 s and 7 = 300 s. Control DNA (undeleted pSV11) is shown in lane C,  $\lambda$  DNA digested with *Ava*II was used as a marker ( $\lambda$ ).

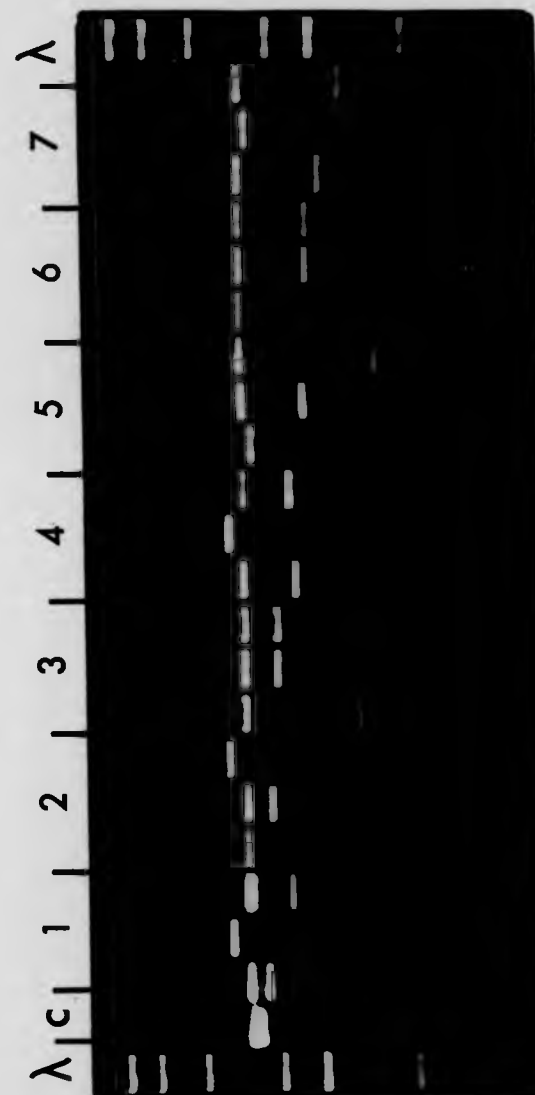
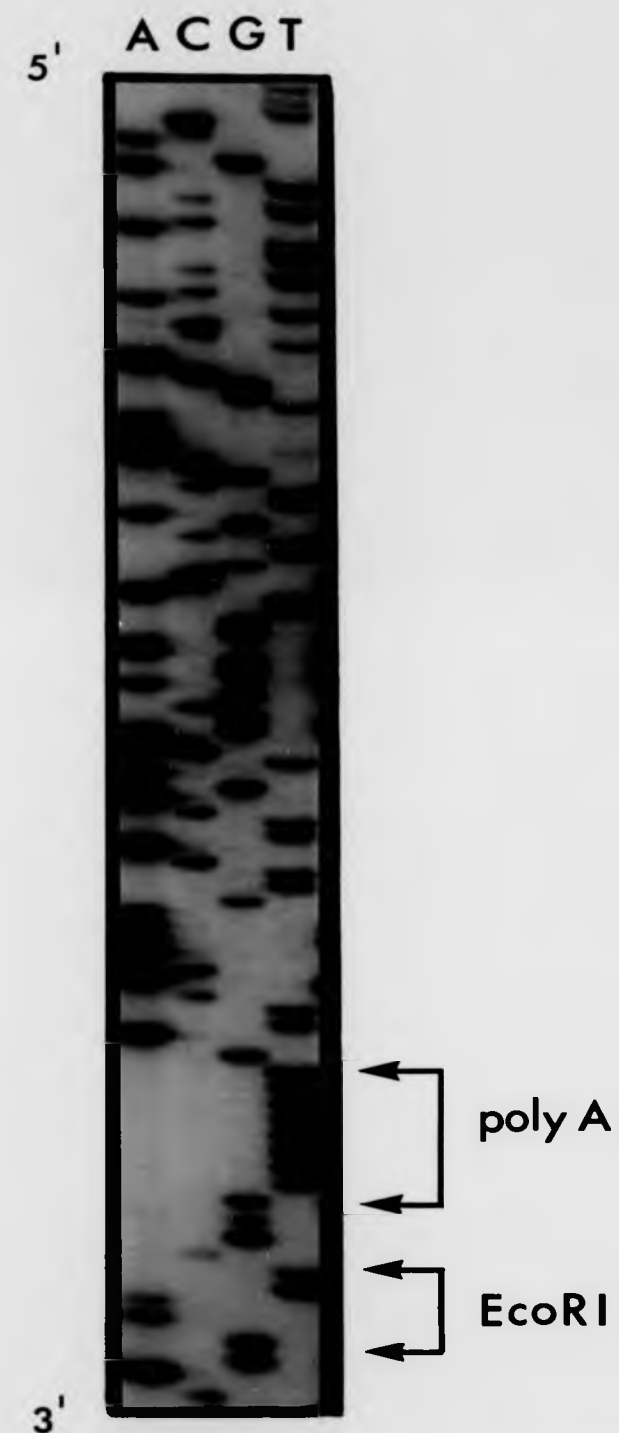


Figure 19. DNA sequence determination of the 3' end containing the poly(A) tail

DNA sequence determination was performed as described in Section 2.2.11 and 2.2.12.1.3 using pSV11 and the T7 primer. The 5' end and 3' ends of the wheat cDNA are indicated as well as the *EcoRI* site in the pUBS vectro used for cloning. Exposure to X-ray film was for 24 h. The different tracks represent the different nucleotides; A (adenine), C (cytosine), G (guanine) and T (thymine) according to IUPAC-IUB (1970). The 5' phosphate (5') and 3' hydroxyl (3') ends of the DNA fragment are indicated.



The *Triticum aestivum* nucleotide sequence (A) and the predicted 545 amino acids (B) of the open reading frame are shown in the one letter code as described by IUPAC-IUB (1970; 1984).

The *Triticum aestivum* nucleotide sequence (A) and the predicted 545 amino acids (B) of the open reading frame are shown in the one letter code as described by IUPAC-IUB (1970; 1984).

1765 TGC GAATCGAACGGTAACCTTCCTCGCCTGATCTGAATGAACCTTTTGTGAATCAAAAAAAAAACC 1828  
-----+-----+-----+-----+-----+-----+-----+-----

polypeptides could be the result of the internal initiation of translation. To investigate this possibility the DNA sequence was analyzed to identify specific nucleotide sequences which are known to be involved in translation initiation.

In most eukaryotes the ribosomal 40S subunit binds at the capped 5' end of the mRNA and scans the sequence until an AUG codon is reached, where it initiates protein synthesis in about 90% of all mRNAs (Kozak, 1983). The wheat cDNA sequence contains an AUG codon starting at position 196, which is surrounded by the nucleotides C (-3); C (-2); A (-1); G (+1) and A (+2) (Fig. 20; Table 5). The coding region in the wheat cDNA fragment is 1629 nucleotides in length and the use of the first AUG at position 196-198 as the initiation triplet would result in the synthesis of a polypeptide with a Mr of 45 661. The use of either the second AUG triplet (position 481) or any of the others as the startpoint of translation would result in the production of polypeptides with a Mr of 36 160 or smaller.

#### 3.2.3.2 Analysis of the 3' non-coding region

In mRNA from eukaryotes the non-coding region varies from 60-300 nucleotides in length and the majority contain a hexanucleotide sequence AAUAAA (Proudfoot & Brownlee, 1976) which is found some 10-30 residues upstream of the polyadenylation site (Fitzgerald & Shenk, 1981). This sequence is considered to be an element directing cleavage and polyadenylation of pre-mRNA. The wheat cDNA sequence contains a 3' non-coding region of 199 nucleotides. Within this region a sequence AAUGAA is found at position 13-18 upstream of the polyadenylation site (Fig. 21).

Berget (1984) suggested that small nuclear ribonucleoproteins (snRNPs) may mediate polyadenylation, by hybridisation to the AAUAAA recognition signal. A second recognition site, CAYUG (Y= pyrimidine), first observed by Benoist *et al.* (1980), is often present either upstream or downstream from the site of poly(A) addition and is thought to direct the precise cleavage point (Berget, 1984). Although exceptions are

Figure 21. The 3' non-coding region of the cDNA from *Triticum aestivum*.

The termination codons in the 199 nucleotide sequence starting with an UGA are underlined. The putative polyadenylation recognition site (boxed) and actual cleavage site (arrow) upstream of the poly(A) region (dotted line) are indicated.

UGAUCGAGCUGUAUGUAAAGCUCUCUUCAGGCCUCCUCCGUUUUUUA  
 GAUGCCUCAGUUCAGGCCAAGCAGUAGAAUUUCGGUGAGAUCAACUG  
 AAAGUGAAAGGUCUAGAGUAAGAGUAGGAUUGCCAUUUUUGCGAAUCG  
 AACGGUAACCUUCCUCGCCUGAUCUGAAUGAACUUUUGUGAAUCAAA  
 ↓  
 AAAAAACC

Table 5. Sequences surrounding the AUG initiation codon.

A. Sequences from plant genes\*

Nucleotide	Upstream							Downstream		
	-4	-3	-2	-1	A	U	G	+1	+2	+3
A	<u>51</u>	<u>62</u>	21	<u>56</u>	<u>100</u>	-	-	5	5	13
G	10	21	5	10	-	-	<u>100</u>	<u>85</u>	15	18
U	13	10	13	7	-	<u>100</u>	-	2	3	15
C	26	7	<u>61</u>	28	-	-	-	8	<u>77</u>	18
Consensus	A	A	C	A	A	U	G	G	C	
Wheat cDNA (1st codon)	C	C	C	A	A	U	G	G	A	A
Wheat cDNA (2nd codon)	G	C	C	A	A	U	G	A	U	U

B. The consensus sequence from animals \*\*

Consensus	C	A	C	C	A	U	G			
-----------	---	---	---	---	---	---	---	--	--	--

\* Data from Lütcke *et al.* (1987).

\*\* Data from Kozak (1984).



known (Kakidani *et al.*, 1982; Furutani *et al.*, 1983; Chin *et al.*, 1984), most animal mRNAs contain only one AAUAAA sequence; however in the few plant RNAs examined multiple recognition sites seem to be the rule rather than the exception (Lycett *et al.*, 1983; Lamb, 1984). The wheat cDNA contains only one copy of the hexameric sequence AAUGAA, while a second recognition site (CAYUG), is not present (Fig. 21). The poly(A) addition site in the wheat sequence is preceded by a C residue (Fig. 21); the significance of this observation is discussed later.

Kohli & Grosjean (1981) examined the context sequences of termination codons of 74 eukaryotic genes (mainly animals and yeast, but no plants). The distribution of the nucleotides in the six positions flanking the termination codon shows that animal and yeast cells strongly prefer purines at the 3' adjacent position, while at the following position pyrimidines are favoured (Table 6). Plant sequences show a similar distribution pattern; however a bias toward the use of a U or A in the third position following the termination codon was observed (Lamb, 1984; Joshi, 1987 and Table 6). The nucleotides surrounding the wheat UGA termination codon i.e. G (-3); U (-2); C (-1); U (+1); C (+2) and G (+3), do not fit the overall plant distribution pattern apart from the Cs at position -1 and +2. The bias towards an A in position -2, found in eukaryotes, is not reflected in the wheat sequence, where a U is observed (Table 6).

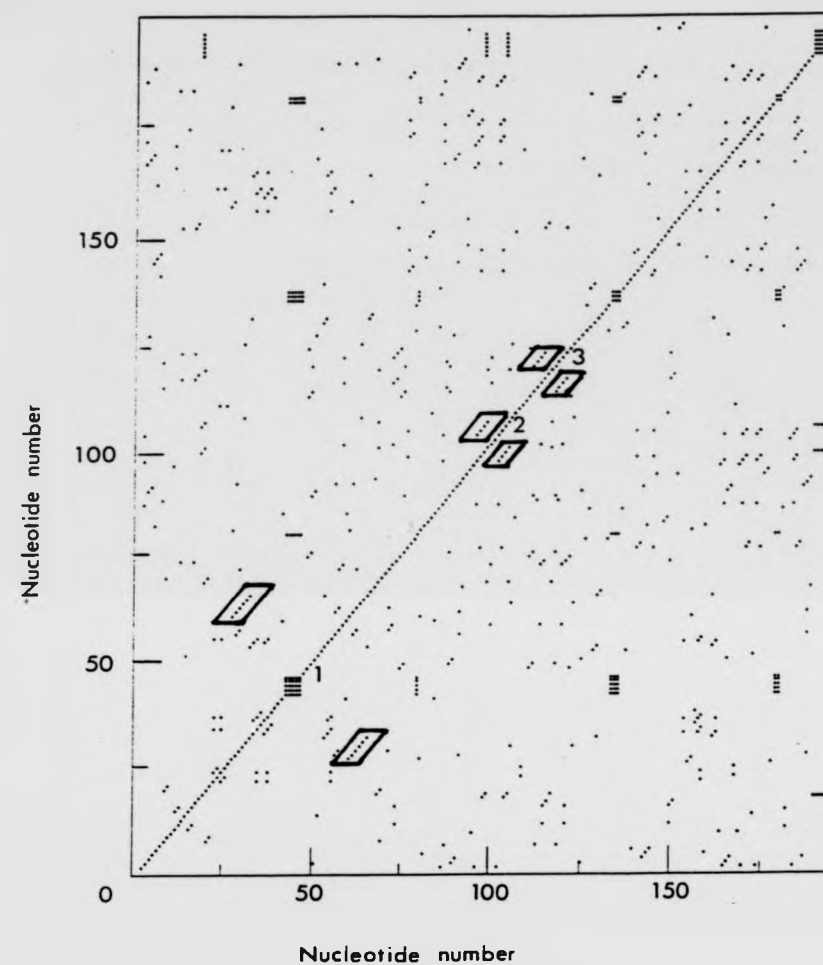
A dot matrix method was used to analyze the 3' non-coding region in a search for direct repeats, palindromes and self-complementary regions such as bulge loops, internal loops and bifunctional loops which all have their own special finger prints on the matrix plot (Maizel & Lenk, 1981; Quigley *et al.*, 1984). When a sequence is compared with itself, repeated regions within the sequence appear as parallel lines. In the wheat sequence three such repeats were identified; the sequence UUCAGGCC starts at position 26 and 59 (Fig. 22A and B, no. 1), whereas two shorter sequences, UGAAAG starting at position 95, and AGAGUA at position 111, are repeated at positions 101 and 117 respectively (Fig. 22A and B, no. 2 and 3). Palindromes, recognizable as diagonal lines in the lower right to upper left direction, are not present in the wheat 3' non-coding sequence (Fig. 22A).



Figure 22. Identification of direct repeats in the *Triticum aestivum* cDNA 3' non-coding region.

Comparison of the 3' non-coding region with itself is presented in the form of a dot matrix. DNA sequences were compared over a moving window of 5 nucleotides, and identity scored if 3 of the 5 nucleotides are identical in any window. The three direct repeats are boxed and in Panel A and the actual nucleotide sequences are shown in Panel B (underlined).

A



B

```

UGAUCGAGCUGUAUGUAAAGCUCUCUUCAGGCCUCCUCCGUUUUUUUA
                                     1a
GAUGCCUCAGUUCAGGCCAAGCAGUAGAAUUUCGGUUGAGAUCAACUG
                                     1b
AAAGUGAAAGGUCUAGAGUAAAGAGUAGGAUUGCCAUUUUUGCGAAUCG
2a  2b  3a  3b
AACGGUAACCUUCCUCGCCUGAUCUGAAUGAACUUUUGUGAAUCAA
AAAAACCC
  
```

Table 6. Preferences and avoidances of nucleotides at positions flanking the natural termination codons.

	Nucleotide upstream of terminator			Codon	Nucleotide downstream of terminator		
	-3	-2	-1		+1	+2	+3
Eukaryotes* (excluding plants, n=74)			69% (G+C)	UAA			
		90% (A)		UGA	82% (G+A)	67% (C+U)	
	68% (C)		0% (C)	UAG			
Plants **			22% (G+C)	UAA			% (C)
		71% (A)		UGA	65% (G+A)	55% (C+U)	52% (U)
	43% (C)		14% (C)	UAG	77% <sup>+</sup>		71% <sup>+</sup> (A+U)
Wheat	G	U	C	UGA	U	C	G

At blank positions no bias is obvious.

\* From Kohli & Grosjean (1981)

\*\* From Lamb (1984).

+ From Joshi (1987).

Comparison of a sequence with its reversed complement reveals regions of self-complementarity that may be involved in the formation of secondary structures in single-stranded molecules (Quigley *et al.*, 1984). Two such regions are found in the wheat cDNA sequence; UUUCGGUUGA, starting at position 78 and its reverse UCAACUGAAA starting at position 90, and the second sequence GUAAGAGUAGG, starting at position 114 of which the complement UGCCAUUUUUGC starts at position 127 (Fig. 23A). The two regions of basepaired nucleotides were joined and the molecules folded as shown in Figure 23B.

#### 3.2.3.3 Codon usage

Only a part of the information present in a genome, varying from 90% for small viruses to only 2% in humans and other mammals, is expressed in the form of proteins. Every amino acid in these proteins, apart from methionine and tryptophan, is represented by several synonymous codons in the DNA which are not utilized with equal frequencies (Grantham *et al.*, 1980). Different genomes from unicellular organisms to higher eukaryotes, differ in their pattern of codon usage (Grantham *et al.*, 1986), e.g. human and yeast nuclear genes and their mitochondria have distinct coding strategies (Grantham *et al.*, 1983). The greatest determinant in creating genetic distances is the G + C content of the degenerate bases those in codon position III, which is the most variable position (Grantham, 1980).

Total human nuclear DNA, like that of other mammals contains about 41% (G + C) (Grantham *et al.*, 1986), whereas all the major families of protein-coding sequences contain a (G + C) content of about 50% (Table 7). A similar observation is made when the % (G + C) of total wheat germ DNA (46% as determined by Marmur & Doty, 1962) is compared with the calculated % (G + C) of known protein coding sequences in monocots, which is 51% (Table 7). The calculated % (G + C) in the coding region of the wheat cDNA is about 50% while the % (G + C) in the non-coding region is significantly

Figure 23. Identification of self-complementary sequences in the *Triticum aestivum* cDNA 3' non-coding region.

The two inverted repeats (underlined) are shown in Panel A. The regions of base-paired nucleotides were joined and the molecules folded (Panel B).

A

```

UGAUCGAGCUGUAUGUAAAGCUCUCUUCAGGCCUCCUCCGUUUUUUUA
GAUGCCUCAGUUCAGGCCAAGCAGUAGAAUUUCGGUUGAGAUCAACUG
AAAGUGAAAGGUCUAGAGUAAGAGUAGGAUUGCCAUUUUUGCGAAUCG
AACGGUAACCUUCCUCGCCUGAUCUGAAUGAACUUUUGUGAAUCAA
AAAAACCC
  
```

B

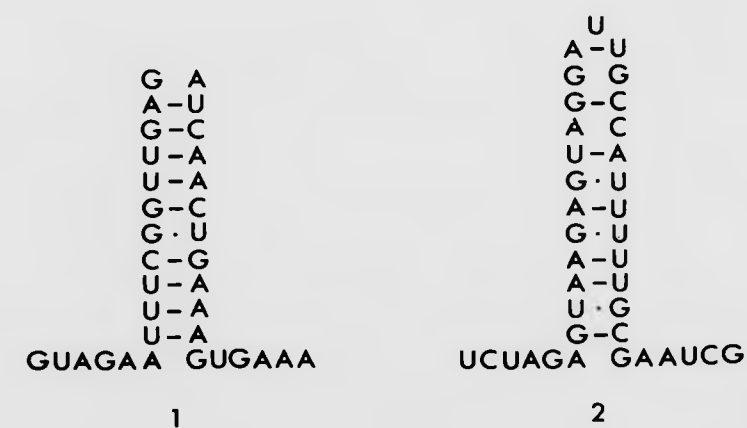


Table 7. Nucleotide composition of pooled protein coding sequences.

	Human	Yeast	<i>E.coli</i>	Plants Dicots	Plants Monocots	Wheat cDNA
Base	n=135	n=64	n=199	n=151	n=53	n=1
	%	%	%	%	%	%
G	26	21	25	23	24	28
C	25	19	23	22	27	24
A	24	30	27	28	26	27
T	25	30	25	27	23	21

Percentages were calculated from data published by Maruyama *et al.* (1986) for human, yeast and *E.coli*, while the data for the plants sequences was published by Murray *et al.* (1989). The number of sequences analyzed is indicated (n).

Table 8. *Triticum aestivum* cDNA nucleotide composition.

Base	Coding region	Non-coding region	Total
	%	%	%
G	28	23	28
C	24	20	24
A	27	26	26
T	21	31	22

Table 9. Coding usage of the *Triticum aestivum* cDNA coding region and of pooled protein encoding sequence from a variety of species.

Amino acid	Codon	Human*	Yeast*	<i>E.coli</i> *	Plants Dicots**	Plants Monocots**	wheat cDNA
		n=135 %	n=64 %	n=199 %	n=151 %	n=53 %	n=1 %
Gly <sup>+</sup>	GGG	22	6	10	12	21	14
Gly <sup>+</sup>	GGA	23	10	6	38	17	12
Gly <sup>+</sup>	GGT	16	69	43	34	18	43
Gly <sup>+</sup>	GGC	39	15	41	16	44	31
Glu	GAG	61	22	29	51	75	69
Glu	GAA	39	78	71	49	25	31
Asp	GAT	42	57	54	58	27	47
Asp	GAC	58	43	46	42	73	53
Val <sup>+</sup>	GTG	48	15	33	29	36	27
Val <sup>+</sup>	GTA	10	13	18	12	8	10
Val <sup>+</sup>	GTT	17	44	32	39	19	29
Val <sup>+</sup>	GTC	25	28	17	20	37	33

		Human*	Yeast*	<i>E.coli</i> *	Plant Dicots**	Plant Monocots**	Wheat cDNA
Ala <sup>+</sup>	GCG	9	7	35	6	22	20
Ala <sup>+</sup>	GCA	18	19	22	25	16	19
Ala <sup>+</sup>	GCT	29	48	20	42	24	28
Ala <sup>+</sup>	GCC	44	26	23	27	38	33
Arg	AGG	23	13	1	25	26	19
Arg	AGA	22	61	2	30	9	29
Ser	AGT	11	13	10	14	8	10
Ser	AGC	27	8	27	18	26	13
Lys	AAG	62	61	24	61	14	68
Lys	AAA	38	39	76	39	25	32
Asn	AAT	42	46	32	45	25	25
Asn	AAC	58	54	68	55	75	75
Met	ATG	100	100	100	100	100	100
Ile	ATA	12	16	4	18	11	8
Ile	ATT	32	51	41	45	24	42
Ile	ATC	56	33	55	37	65	50



		Human*	Yeast*	<i>E.coli</i> *	Plants dicots**	Plants monocots**	Wheat cDNA
Thr <sup>+</sup>	ACG	9	8	21	8	21	8
Thr <sup>+</sup>	ACA	24	22	9	27	14	33
Thr <sup>+</sup>	ACT	24	40	21	35	19	33
Thr <sup>+</sup>	ACC	43	30	49	30	46	25
Trp	TGG	100	100	100	100	100	100
End	TGA	52	48	22	33	34	100
Cys	TGT	40	72	42	44	30	33
Cys	TGC	60	28	58	56	70	67
End	TAG	15	29	7	19	36	0
End	TAA	33	23	71	48	30	0
Tyr	TAT	40	42	48	43	21	50
Tyr	TAC	60	58	52	57	79	50
Leu	TTG	11	44	10	26	14	14
Leu	TTA	6	23	9	10	3	0
Phe	TTT	40	48	44	45	25	57
Phe	TTC	60	52	56	55	75	43

		Human*	Yeast*	<i>E.coli</i> *	Plants dicots**	Plants monocots**	Wheat cDNA
Ser <sup>+</sup>	TCG	6	7	13	6	14	17
Ser <sup>+</sup>	TCA	12	15	9	19	11	10
Ser <sup>+</sup>	TCT	19	36	21	25	15	33
Ser <sup>+</sup>	TCC	25	21	20	18	26	17
Arg <sup>+</sup>	CGG	15	2	5	5	13	14
Arg <sup>+</sup>	CGA	10	4	4	8	4	14
Arg <sup>+</sup>	CGT	7	17	50	21	12	14
Arg <sup>+</sup>	CGC	23	3	38	11	36	10
Gln	CAG	73	23	72	41	46	78
Gln	CAA	26	77	28	59	54	22
His	CAT	41	52	44	54	33	33
His	CAC	59	48	56	46	67	67
Leu <sup>+</sup>	CTG	44	8	61	9	28	25
Leu <sup>+</sup>	CTA	6	12	2	8	9	7
Leu <sup>+</sup>	CTT	11	9	8	28	15	34
Leu <sup>+</sup>	CTC	22	4	10	19	31	20

		Human*	Yeast*	<i>E.coli</i> *	Plants Dicots**	Plants Monocots**	Wheat cDNA
Pro <sup>+</sup>	CCG	2	10	62	9	23	17
Pro <sup>+</sup>	CCA	23	56	18	42	34	28
Pro <sup>+</sup>	CCT	28	27	13	32	17	22
Pro <sup>+</sup>	CCC	37	7	7	17	26	33

\* data from Maruyama *et al.* (1986).

\*\* data from Murray *et al.* (1989).

Table 10. Percentages (G+C) content in protein encoded sequences from different species.

A. Percentages (G+C) content in pooled mRNA sequences.

	Human*	Yeast*	<i>E.coli</i> *	Plants Dicots n=151	Plants Monocots n=53	<sup>2</sup> 6P Wheat cDNA n=1
% (G+C) total	51	40	48	45	51	52
% (G+C)qIII	52	25	56	30	59	44
% (G+C)III	58	38	53	44	64	53

The number of sequences analysed is indicated (n).

\* percentages calculated from data published by Maruyama *et al.* (1986).

\*\* percentages calculated from data published by Murray *et al.* (1989).

B. Percentage (G+C) content in individual wheat sequences

	cDNA	CAB*	SS*
% (G+C) total	52	55	54
% (G+C)qIII	44	65	70
% (G+C)III	53	75	82

CAB, chlorophyl A/B binding protein mRNA

SS, Rubisco small subunit mRNA

\* percentages calculated from data published by Maruyama *et al.* (1986).

lower, i.e. 43% (Table 8). The % (G + C) of the complete cDNA sequence is 52% (Table 8).

The codon strategy used by a particular species becomes evident when codon frequencies of protein coding regions are compared. To eliminate the influence of amino acid frequency on codon frequency, only the eight sets of quartet codons; i.e. codons in which the change of the third base does not affect the amino acid it codes for (Grantham, 1980), were used to determine the % (G + C) III (Table 9, marked +). The low percentage, (G + C)qIII = 44%, calculated for the wheat sequence suggest a preference for the use of an A or T rather than a G or C residue, and does not conform to the % (G + C)qIII calculated for all known monocot sequences (Table 10A). Among the quartet codons in the wheat sequence only Val shows a bias towards G or C in the third position [% (G + C) = 60, Table 9]. However Thr displays a significant preference for an A or a T in the third position [% (G + C) = 33]. The duet codons with a preference for G or C in the third position [% (G + C) > 60%] are Glu, Lys, Asn, Lys, Gln and His, while no bias towards a particular nucleotide is observed for the amino acids Asp and Tyr (Table 9). The % (G + C) in the third position of all degenerate codons of the wheat cDNA coding-region is 52% (Table 10A). These observations are discussed later (Section 3.2.6.3).

#### 3.2.4 IDENTIFICATION OF THE TRANSLATED POLYPEPTIDE

The Rubisco subunit binding protein occurs as an oligomer of two types of polypeptide, named  $\alpha$  and  $\beta$ , with apparent *M<sub>r</sub>* of about 61 000 and 60 000 respectively in the case of *Pisum sativum* (Hemmingsen & Ellis, 1986). Antibodies raised against the purified oligomer from pea recognize similarly sized polypeptides from wheat (Fig. 5 and 6) and were used to isolated the wheat cDNA sequence (Section 3.1.3). In order to establish the identity of the isolated wheat cDNA sequence the predicted aminoterminal sequence was compared with the determined aminoterminal of the purified mature  $\alpha$  subunit of the Rubisco subunit binding protein from wheat and pea

(Fig. 24) The predicted sequence corresponds well with the determined  $\alpha$  subunit from wheat; i.e. 17 matches out of 20, and pea; i.e. 13 matches out of 20. There is less correspondence between the predicted wheat sequence and the aminoterminal sequence of the purified mature pea  $\beta$  subunit; i.e. 9 matches out of 20. Four of these matches were seen only when extra residues were added to the pea  $\beta$  subunit in order to optimize the alignment. It was concluded that the isolated cDNA fragment encodes the entire mature Rubisco subunit binding protein  $\alpha$  subunit plus two amino acid residues of the presequence. The mature  $\alpha$  subunit from wheat contains 543 amino acids and has a calculated Mr of 57 393.

### **3.2.5 ANALYSIS OF THE AMINO ACID SEQUENCE OF THE RUBISCO SUBUNIT BINDING PROTEIN ALPHA SUBUNIT.**

#### **3.2.5.1 Analysis of the amino acid composition**

All proteins tend to have similar amino acid compositions regardless of their origin; however prokaryotic sequences contain significantly less cysteine (0.9%) than those from eukaryotes especially human sequences (3.0%) (Doolittle, 1986). The amino acid composition of the wheat  $\alpha$  subunit is remarkably similar to that reported for *E.coli* sequences which show a high percentage of alanine and a low percentage of cysteine (Table 11). This similarity might reflect a prokaryotic origin of the binding protein  $\alpha$  subunit.

#### **3.2.5.2 Determination of the hydropathy profile**

The structure of a globular protein is largely determined by its polar and apolar residues that are respectively on the surface in contact with the aqueous medium, and buried in the interior of the protein molecule. The hydropathy (= feeling for water) index, determined for each of the 20 amino acids, reflects the relative hydrophobicity

Figure 24. Comparison of the aminoterminal sequences of the Rubisco subunit binding protein subunits from *Triticum aestivum* and *Pisum sativum*.

A lyophilized sample of the *Triticum aestivum* binding protein  $\alpha$  subunit (kindly provided by A.J. Keys) was subjected to automated solid-phase Edman degradation by the SERC protein sequencing unit at the Department of Biochemistry, University of Leeds. The determined  $\alpha$  and  $\beta$  subunits from *Pisum sativum* have been published (Musgrove *et al.*, 1987). Asterisk indicate identical residues; dots indicate gaps to maximize the alignment.

Wheat predicted	GADAKEIAFDQKSR AALQAGVE
Wheat determined $\alpha$	* * * * * X A P E I A F D Q Q S R A A L Q A G V E

Wheat predicted	GADAKEIAFDQKSR AALQAGVE
Pea determined $\alpha$	* * * * * A A K D I A F D Q H S R S A M Q A G I D

Wheat predicted	GADAKEIAFDQKSR A A - LQAGV
Pea determined $\beta$	* * * * * A K E L H F N K D G S A I R K L Q N G V

Table 11. Amino acid compositions of the mature *Triticum aestivum*  $\alpha$  subunit and pooled protein sequences from different species.

Amino acid residue	Mature $\alpha$ subunit		<i>E.coli</i> *	Eukaryotes*	All*
	number of residue	%	%	%	%
Gly	42	7.7	7.5	7.6	7.2
Pro	18	3.3	3.9	4.9	5.2
• Asp	30	5.5	5.6	5.4	5.3
• Glu	51	9.4	6.6	6.8	6.3
+ Ala	69	12.7	10.1	7.2	7.8
• Asn	20	3.7	4.1	4.1	4.3
• Gln	18	3.3	4.3	4.3	4.2
Ser	30	5.5	5.5	7.0	6.8
Thr	36	6.7	5.3	5.6	5.9
• Lys	38	7.0	5.7	6.5	5.9
• Arg	21	3.9	5.9	4.9	5.1
• His	3	0.5	2.0	2.4	2.3
+ Val	48	8.9	7.5	6.8	6.6
+ Ile	50	9.2	5.8	5.0	5.3
+ Met	6	1.1	2.6	2.0	2.3
+ Cys	3	0.5	0.9	1.7	1.9
+ Leu	44	8.1	9.5	8.7	9.1
+ Phe	7	1.3	3.5	4.1	3.9
Tyr	8	1.5	2.7	3.8	3.2
Trp	1	0.2	1.0	1.4	1.4

\* Data from Doolittle (1986).

Hydrophilic (•) and hydrophobic (+) amino acids are indicated.



and hydrophilicity and as such is a measure of polarity (Kyte & Doolittle, 1982). Whether a particular amino acid ends up either inside or on the surface of the folded protein molecule depends on its own hydrophathy value as well as that of its surrounding neighbours. The continuous determination of the average hydrophathy index of a segment of amino acids that moves through the Rubisco subunit binding protein  $\alpha$  subunit sequence from the amino to the carboxyterminus produces a graphic display of the hydrophathy character (Fig. 25). A hydrophobic region is present in the middle of the sequence while the carboxyterminus contains more hydrophilic amino acids. In the  $\alpha$  subunit sequence 42% of the amino acids were assigned as hydrophobic (Table 11, amino acids labelled +) whereas only 32% display a hydrophilic character (Table 11, amino acids labelled •; Kyte & Doolittle 1982).

#### 3.2.5.3 Secondary structure prediction

The information to fold a protein into its secondary structure i.e. the local arrangement of the polypeptide backbone, resides within the primary amino acid sequence. In a globular soluble protein four secondary structure conformations are recognized: helices, sheets, turns and random coils. The observation that each amino acid differs in its ability to form a particular secondary structure was exploited by Chou & Fasman (1978) who used proteins with known X-ray structures to calculate the conformational parameter, i.e. the tendency for an amino acid to appear in a particular conformation, for each of the 20 amino acids. Individual amino acids are assigned as e.g. strong helix formers, indifferent helix formers, weak sheet formers etc. and this formed the basis for the identification of secondary structures in a polypeptide chain. The prediction of the secondary structures in the Rubisco subunit binding protein  $\alpha$  subunit amino acid chain is shown in Figure 26. At the aminoterminal, the polypeptide chain starts with an  $\alpha$ -helix followed by a region of  $\beta$ -strands and turn; an arrangement which is also found in several other areas in the chain. Segments of  $\alpha$ -helices and  $\beta$ -sheets are generally short as they are limited to the diameter of the globular protein.

Figure 25. Hydropathy profile of the Rubisco subunit binding protein  $\alpha$  subunit.

The hydropathy index, determined over a window of 21 residues, is plotted against the amino acid sequence. The long hydrophobic region in the middle of the protein is indicated (————). The method of Kyte and Doolittle (1982) was used.

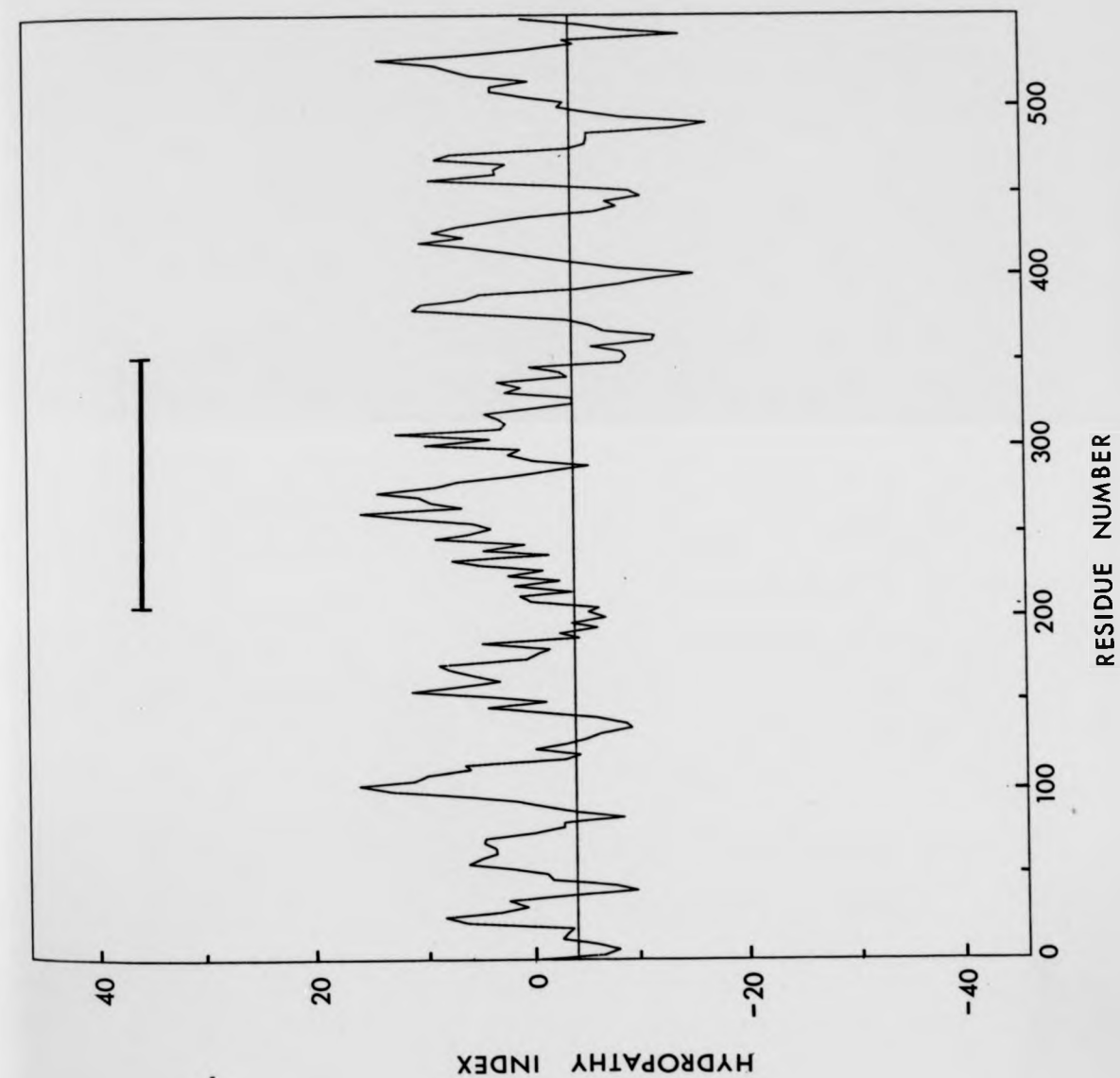
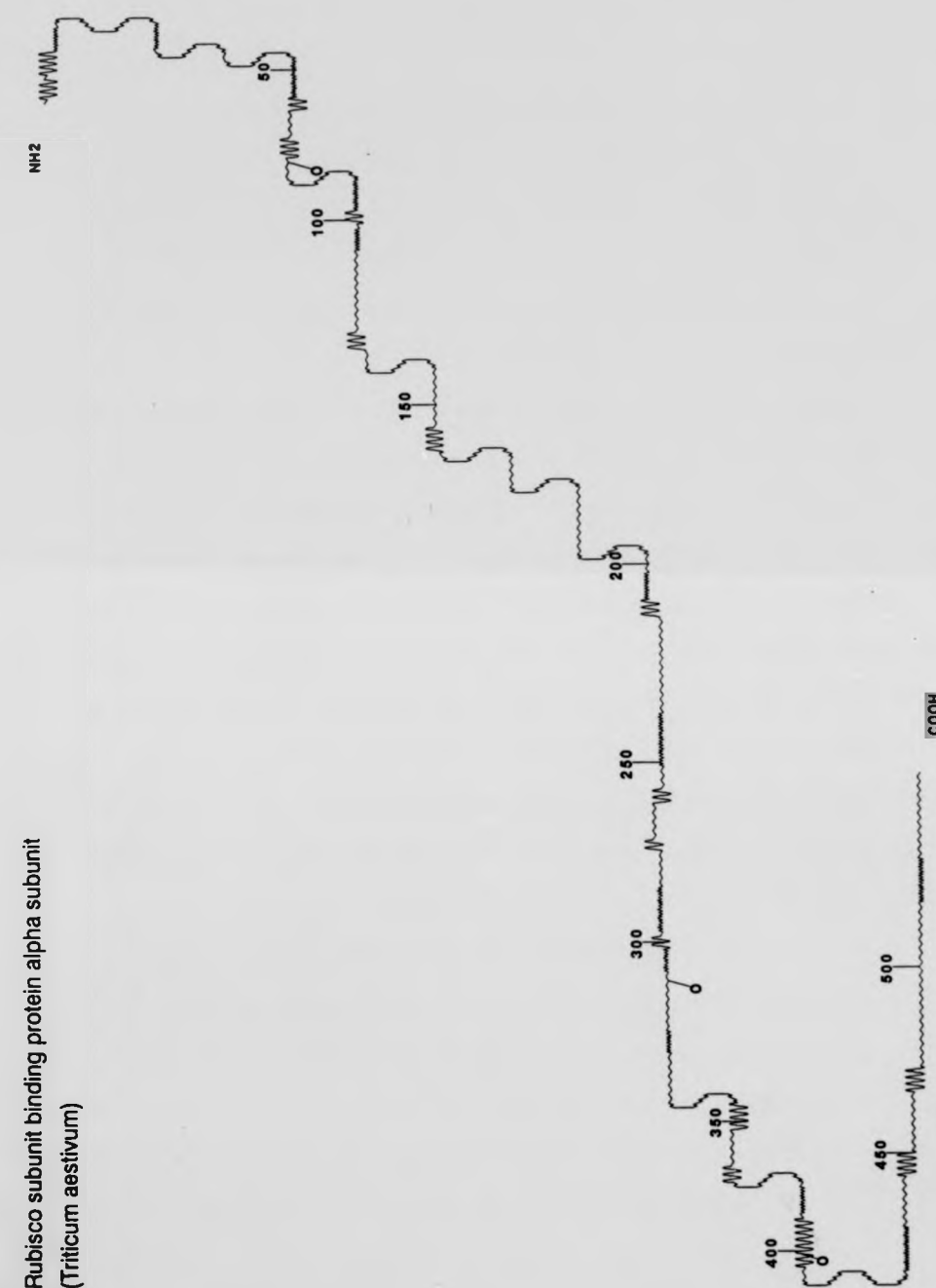


Figure 26. Secondary structure predictions of the Rubisco subunit binding protein  $\alpha$  subunit.

The secondary structure prediction (Chou & Fasman, 1978) is presented as a two dimensional plot with the aid of the 'plotstructure' program from the GCG software package (Devereux *et al.*, 1984). Helices are shown with a sine wave,  $\beta$  sheets with a sharp saw tooth wave, turns with 90 degree turns, and coils with a dull saw tooth wave. Putative glycosylation sites are indicated (—O).

The quality i.e. the percentage of residues placed in the correct structural class ( $\alpha$ -helix,  $\beta$ -sheet, turn or random coil), of a secondary structure prediction, such as presented in this figure is approximately 45%.



The length of an  $\alpha$ -helix is between 10 and 15 residues while that of a  $\beta$ -strand is 3 to 10 residues (Creighton, 1984). The folding types that were identified in the  $\alpha$  subunit were: a strong  $\alpha$ -helix (around position 400), and  $\alpha/\beta$  conformations (around positions 10, 50, 100, 200, 300, and 400), whereas the last 75 amino acids largely fold into a random coil conformation.

#### 3.2.5.4 Identification of possible active sites and consensus binding sites

To obtain more information about the possible function of the Rubisco subunit binding protein  $\alpha$  subunit, the amino acid sequence was searched for known active site and consensus binding site sequences which are listed in Table 12 (Doolittle, 1986). Neither any of the active sites nor a zinc or cytochrome *c* thioester binding site were identified. However, the sequence N X T/S (X = various residues), known to be required for glycosylation, is found at positions 81-83, 308-311 and 401-403 in the mature  $\alpha$  subunit (N D S and N A T (twice) respectively). These sites are marked (--O) as possible glycosylation sites in the secondary structure prediction graph (Fig. 26).

Recently a consensus sequence of 12 amino acid residues, representing a calcium binding domain has been published i.e. E L X L L X X L \* X \* X \* G X L \* X X \* L X X L L X X L in which \* indicates an oxygen-containing residue (D, E, N, Q, S, T), L indicates a hydrophobic residue (L, V, I, F, M), G is glycine and X is any amino acid (Wales *et al.*, 1989). Although the glycine at position 256 in the mature  $\alpha$  subunit is surrounded by amino acids out of which 8 match the required 15 residues for a calcium binding site, this is not considered to be sufficient to identify this region as a putative site. The  $\alpha$  subunit contains three cysteine residues which potentially could be involved in the formation of disulphide bonds, but there is not evidence that the binding protein contains such bonds.

The Rubisco subunit binding protein oligomer from pea is known to dissociate reversible *in vitro* in the presence of MgATP (Bloom *et al.*, 1983; Hemmingsen & Ellis, 1986; Musgrove & Ellis, 1986; Musgrove *et al.*, 1987), while the purified pea protein

Table 12. Active sites and consensus binding site sequences.

Enzyme active sites\*

---

Common serine proteases	G D S G G
Subtilisin-type serine proteases	T M S A
Sulphydryl proteases	G X C W/Y
Acid proteases	U D T G
Hydroxy-kinases	D F G (X) <sub>n</sub> A P E
Guanidine-kinases	T C P (X) <sub>n</sub> N L G T
Phosphoglyceraldehyde dehydrogenase	C T T N C
Triose phosphate isomerase	U A Y E P
Phosphoglucomutase	T A S H D
Phospholipase A	C C X X H D
Lactamase	S X X K

---

Binding sites\*

---

Asn-linked carbohydrate	N X S/T
Zinc-binding	C/H (X) <sub>n</sub> C/H (X) <sub>n</sub> C/H (X) <sub>n</sub> C/H
Cell adhesion	R G D
Cytochrome c thioester	C X X C H

---

\* From Doolittle (1986).

has been reported to show ATPase activity (Chaudhari *et al.*, 1987), indicating the presence of an ATP binding site. Walker *et al.* (1982) were the first to identify two regions of highly conserved amino acids in ATP-requiring enzymes: segment A which contains the sequence G X X X X G K T/S X X X X X X I/V, and segment B containing the common sequence R/K X X X G X X X L \* \* \* D, where X = various residues and \* = hydrophobic amino acids. Although the F1ATPase subunits initially studied by Walker *et al.* (1982) contain both the segments A and B, in other proteins only one of the two sequences is present e.g. myosin contains segment A but not segment B, while in phosphofructokinase only the sequence of fragment B is observed. The mature wheat  $\alpha$  subunit sequence was scanned in a search to identify regions similar to the reported segments A and B but without any success.

Wieringa *et al.* (1986) reported a "fingerprint" sequence, which is a set of rules describing the type of amino acid that should occur at 11 specific positions in a peptide fragment with ADP-binding properties. They showed that this "fingerprint" can be used to identify nucleotide binding proteins. In the Rubisco subunit binding protein  $\alpha$  subunit sequence no match is found to this "fingerprint" when comparisons are carried out in the conventional way i.e. by scanning from the aminotermminus to the carboxyterminus. However, a good match is seen if the comparison is made in the opposite direction, i.e. from the carboxyterminus to the aminotermminus. Thus histidine 419, alanine 416, glycine 414, 412 and 408, alanine 403 and 404, and glutamic acid 386 match the "fingerprint" provided that one extra residue is inserted between glycine 412 and 408, and between glycine 408 and alanine 403 (Fig. 27). This remarkable match will be referred to in the discussion.

Another category of binding sites are those recognized by antibodies, most of which are discontinuous i.e. consist of residues that are apart in the sequence, rather than continuous. The small number of sites that are continuous often correspond to loops and/or protruding regions on the surface of the protein (Barlow *et al.*, 1986), and obviously properties such as hydrophilicity (Hopp & Woods, 1981), accessibility (Novotny *et al.*, 1986), mobility (Westof *et al.*, 1984) and protrusion (Thornton *et al.*,





1986) can be used to predict which parts of the polypeptide chain are likely to contain the potential antigenic peptides. Historically, antigenic predictions have utilized a single approach e.g. either hydropathy or surface flexibility. Recently Jameson and Wolf (1988) have described the antigenic index, a measure of potentially exposed surface regions of a protein, which is defined by an algorithm that integrates flexibility components like hydrophilicity, surface probability, backbone flexibility and secondary structure. This algorithm was utilized to predict potential antigenic regions in the binding protein  $\alpha$  subunit sequence. The aminoterminal end of the polypeptide is the most potential antigenic region (Fig. 28). Although this region (between residue 5 and 50) is not particularly hydrophilic (Fig. 25), it is probably very flexible as turns and  $\beta$ -strands have been identified in this area (Fig. 28); thus this region is likely to appear on the surface of the  $\alpha$  subunit. Other possible antigenic sites were identified around positions 80, 140, 200, 240 and 350, whereas the carboxyterminus from position 350-543 (end) is not likely to act as an antigenic region.

### 3.2.6 DISCUSSION

#### 3.2.6.1 Identification of the *Triticum aestivum* cDNA fragment

The isolated wheat cDNA fragment contains an open reading frame that codes for a polypeptide with a Mr of 57 521. This polypeptide represents the entire mature Rubisco subunit binding protein  $\alpha$  subunit plus two amino acid residues of the presequence (Fig. 24). Comparison of the predicted wheat sequence with the determined aminoterminal sequence of the purified wheat binding protein  $\alpha$  subunit shows three discrepancies which may be the result of incomplete recoveries of aminoterminal and lysine residues associated with the sequencing method used. Alternatively, they may reflect varietal differences, as the  $\alpha$  subunit was obtained from variety 'Fico' while the cDNA was isolated from variety 'Chinese Spring'. The observation that the aminoterminal sequences of the  $\alpha$  and  $\beta$  subunit of the binding protein from pea show

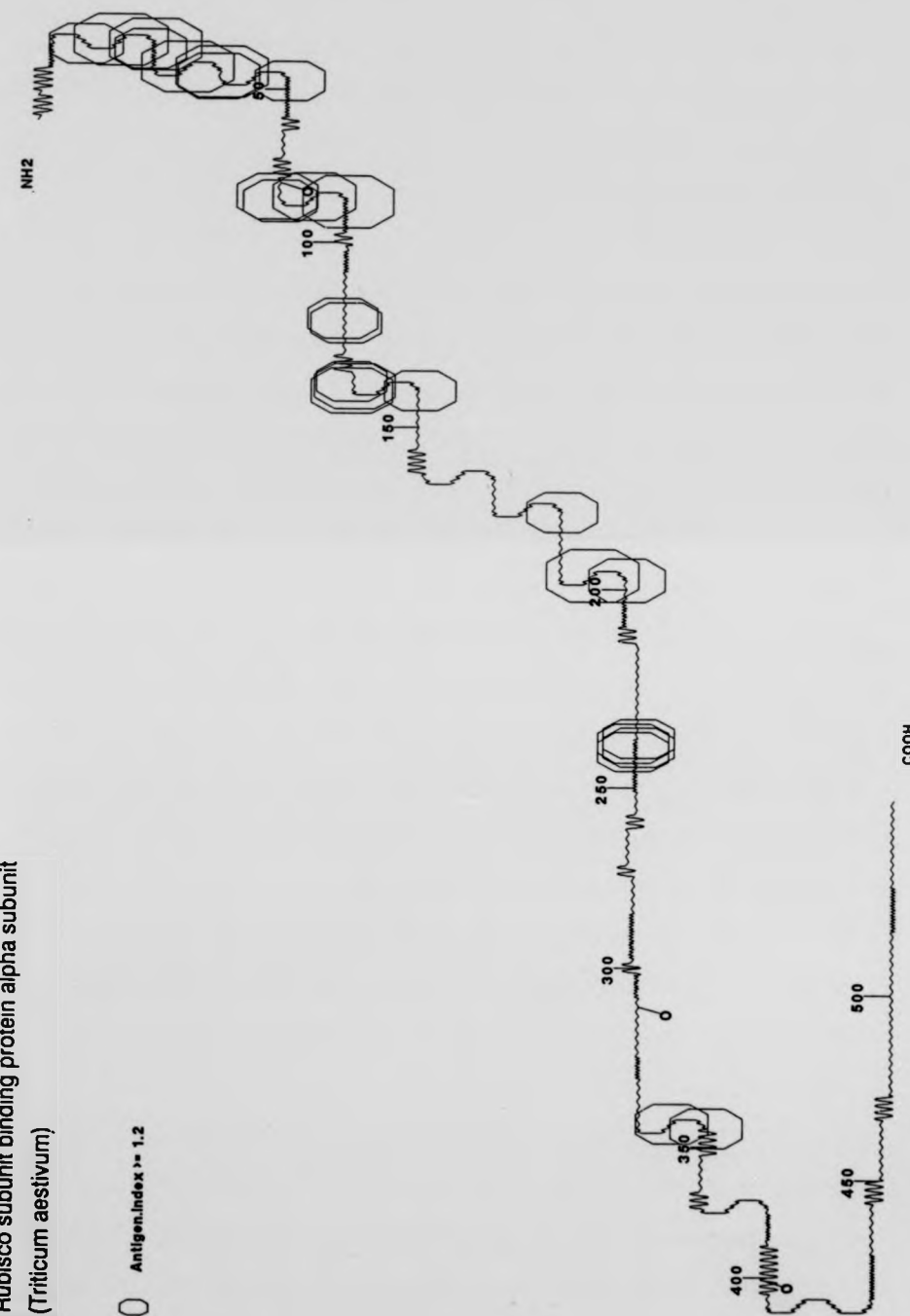


Figure 28. Antigenic profile of the Rubisco subunit binding protein  $\alpha$  subunit.

The antigenic index is determined using the algorithm described by Jameson and Wolf (1988) and is a measure for potential exposed surface regions. The antigenic index is superimposed over the secondary structure prediction with a special symbol, the size of which is proportional to the value of the attribute (Devereux *et al.*, 1984).

Rubisco subunit binding protein alpha subunit  
(*Triticum aestivum*)

○ Antigen.Index  $\propto$  1.2



amino acid similarity led Ellis *et al.* (1987) to suggest that the two subunits are related. The occurrence of limited identity between the wheat sequence and the mature  $\beta$  subunit sequence from pea indicates that the wheat  $\alpha$  and the pea  $\beta$  subunits are related (Fig. 24).

Translation of the wheat cDNA transcript *in vitro* results in the synthesis of two polypeptides with apparent  $M_r$  of 59 000 and 55 000 (Fig. 18). A eukaryotic mRNA normally possesses a capped 5' end which stabilizes the transcript and enhances the binding of the 40S ribosomal subunit, and a downstream AUG codon where translation is initiated. In the cell-free system used the synthesized mRNA is not expected to be capped; however translation *in vitro* using a wheat germ extract is not affected by lack of capping (R.W. Old and M.J May, personal communication). Certain nucleotides surrounding the AUG codon influence the ability of an AUG triplet to be recognized as an initiation codon (Kozak, 1984) and these consensus nucleotides are different in animal and in plant sequences (Table 5; Lütcke *et al.*, 1987). Since eukaryotic ribosomes are not too demanding when it comes to choosing the initiation codon, and even in the absence of a long consensus signal the start site is known to be the most 5' AUG in 90% of cases, the AUG at position 196 in the wheat cDNA transcript is the most likely to be used as the initiation codon. The polypeptide derived by initiation at this position has a calculated  $M_r$  of 45 661, but migrates on an SDS polyacrylamide gel at a position indicating an apparent  $M_r$  of 59 000. A similar discrepancy is observed for the mature wheat binding protein  $\alpha$  subunit which has a calculated  $M_r$  of 52 393, but which migrates in an SDS polyacrylamide gel at an apparent  $M_r$  of 62 000 (Fig. 9 and Fig. 10). A second possibility could be that no AUG initiation codon is used and the entire transcript is translated *in vitro*. The entire transcript contains 10 additional nucleotides (CCCGCUUAAG) upstream of the start of the mature  $\alpha$  subunit coding region (Fig. 20). If translation starts at position 1 in the transcript a very small peptide is synthesized since the third triplet is the translational termination codon UAA. It is concluded that translation *in vitro* starting at position 1 can not explain the synthesized polypeptide with an apparent  $M_r$  of 59 000. The second translation product of the wheat

cDNA transcript, with an apparent Mr of 55 000 (Fig. 18) is either the result of the internal initiation of translation using the AUG codon at position 481, which would produce a polypeptide with a calculated Mr of 361 36 or the result of degradation of the first polypeptide. Attempts to immunoprecipitate the synthesized polypeptides with the antibodies raised against the purified binding protein from *P.sativum* were unsuccessful.

#### 3.2.6.2 The *Triticum aestivum* binding protein $\alpha$ subunit amino acid sequence

The mature  $\alpha$  subunit is a slightly hydrophobic protein with a calculated Mr of 52 393 and an amino acid composition that is remarkably similar to that reported for *E.coli* proteins, (Table 11) which suggests a possible prokaryotic origin. The amino acid sequence of the mature  $\alpha$  subunit contains three possible glycosylation sites represented by the tripeptide Asn X Thr/Ser. Either the Asn residue or the Thr/Ser amino acid can be involved in the formation of the carbohydrate-peptide bond i.e. N-glycosylation and O-glycosylation respectively (Aubert *et al.*, 1976). It is the secondary structure of the domain in which the tripeptide is located, as well as the presence of certain nearby amino acids, e.g. proline in the case of O-glycosylation, that determine whether an Asn X Thr/Ser site is used, and for which mode of glycosylation. It seems unlikely that the putative sites in the  $\alpha$  subunit are in fact glycosylated, since the enzymes that carry out glycosylation are believed to occur only within the endoplasmic reticulum (Hubbard & Ivatt, 1981). The  $\alpha$  subunit is part of the Rubisco subunit binding protein oligomer which is located in the chloroplasts of plants and has not yet been reported to be glycosylated.

The most interesting finding to emerge from the analysis of the deduced amino acid sequence of the wheat  $\alpha$  subunit is that a possible dinucleotide binding site is present between positions 386 and 419 (Fig. 27). This region shows 9 out of the 11 matches reported for the nucleotide binding site fingerprint (Wieringa *et al.*, 1986) but

only when the sequence is read from the carboxy to the aminotermminus and an extra amino acid is allowed between glycine (408) and glycine (412), known as the 'core' of the fingerprint, and between residues 403 and 400. The length of the loop in this fingerprint is reported to be variable (Wieringa *et al.*, 1986), and a better match of 10 out of the 11 positions characteristic for the fingerprint is seen when only 3 amino acid residues form the loop structure [Fig. 27, (b)]. The core ( G X G X X G, where X= various residues) is highly conserved among nucleotide binding proteins; however some flexibility seems to be allowed, as in yeast alcohol dehydrogenase, with only 7 out of the 11 matches, one of the essential glycine residues in the core is changed to an alanine (Wieringa *et al.*, 1985). A similar change i.e. an alanine substituted for a glycine has also been reported for the cytoplasmic porcine malate dehydrogenase (Birktoft *et al.*, 1982). If a loop of 5 amino acids is allowed in the putative nucleotide binding site of the  $\alpha$  subunit, an acidic amino acid, glutamic acid (position 388 in the wheat sequence and position 30 in the fingerprint), is found rather than the small hydrophobic residue predicted from the fingerprint. Such a change has also been reported for the p21 protein (gene product of H-ras-1, human) a protein with GTPase activity (McGrath *et al.*, 1984); and modelbuilding studies have shown that the peptide in the p21 protein can be built as an ADP-binding site (Wieringa & Hol, 1983).

It has been reported that the Rubisco subunit binding protein purified from *P.sativum* shows both weak ATPase activity (Chaudhari *et al.*, 1987) and the ability to be dissociated into monomers by incubation with MgATP (Musgrove *et al.*, 1986; Bloom *et al.*, 1983). The identification of a possible, dinucleotide binding site in the wheat  $\alpha$  subunit is thus not unexpected. What is remarkable however, is that this site is apparent only when the sequence is read from the carboxy to the aminotermminus. No precedent for such a reversed sequence is known and this observation raises two fundamental questions that demand further study. Firstly, can similar tertiary structures be constructed with the nucleotide fingerprint in either direction, and secondly, should database searches in the future be run in both directions to see whether other cases of reverse sequences occur? The importance of determining the tertiary structure in order

to allocate a functional ATP-binding site was demonstrated by Fry *et al.* (1986), who determined the ATP-binding site in adenylate kinase by NMR spectroscopy. Despite the little primary sequence homology between adenylate kinase and dehydrogenases (Wieringa *et al.*, 1985), Fry *et al.* showed similarity in three-dimensional structure and function between their nucleotide binding sites, in which the functional amino acid residues often come from regions that differ in their protein sequences. These observations suggest that there is likely more than one way in which a protein can bind a dinucleotide molecule and, probably more than one way to fold a nucleotide binding site. It may well be that the differences between the fingerprint and the reversed sequence in the binding protein  $\alpha$  subunit are required to accommodate the fact that the peptide runs in the reverse direction. More information as to whether the reversed sequence is present in other proteins and which amino acids are directly involved in the binding of ATP, is required to establish whether the identified region in the  $\alpha$  subunit sequence is really a functional ATP binding site (see Section 3.3.6).

#### 3.2.6.3 The *Triticum aestivum* binding protein $\alpha$ subunit nucleotide sequence

The wheat  $\alpha$  subunit mRNA contains a 3' non-coding region of 199 nucleotides and starts with the termination triplet UGA. The distribution of termination codons in nuclear genes from plants shows that UAA is most used in dicots, while the UAA and UGA codons are equally used in monocots, 35% and 37% respectively (Table 13). In contrast, in other eukaryotes, the most widely used codon is UAA (44%) followed by UGA (39%) and UAG (17%).

The hexanucleotide sequence AAUGAA starting at position 1799 in the wheat mRNA is most likely the recognition element for small nuclear ribonucleoproteins involved in polyadenylation. Although the AAUAAA is the most frequent sequence, nine other sequences containing single point mutations have been found in nature at appropriate distance from the polyadenylation site, and all are considered functional

Table 13. Distribution of termination codons.

Codon	Plants*	Dicots*	Monocots*	Monocots* no storage proteins	Other eukaryotes**
	n=204	n=151	n=53	n=39	n=351
	%	%	%	%	%
UAA	43	48	30	35	44
UGA	33	33	34	37	39
UAG	24	19	36	28	17

The number of sequences analyzed are indicated (n)

\* Data taken from Murray *et al.*, 1989.

\*\* Other eukaryotes include human, hamster, mouse, rat, bovine, chicken, anglerfish, flounder, eel, trout, carp and salmon (Maruyama *et al.*, 1989).

hexamers (Birnstiel *et al.*, 1985). The AAUGAA sequence found in the wheat mRNA is thought to support efficient processing. Other wheat sequences have been reported to contain variations of the AAUAAA motif, e.g. Rubisco small subunit, UAAACUA; gliadin  $\alpha/\beta$ , UAGAAC, gliadin  $\gamma$ , UGAAAA and glutenin  $\lambda c11$ , UGAUAG, and all are sufficient for correct 3' RNA processing in wheat (Joshi, 1987). An unaltered AAUAAA motif is present in only 39% of the examined plant genes whereas one base substitution occurs in 54% of these sequences (Lamb, 1984; Joshi, 1987). When SV40 AATAAA mutant DNAs were injected into *Xenopus* oocytes, reduced cleavage efficiency is observed. The cleavage efficiency for the AATATA mutant was reported to be 10% of that of the wild type, while the AACAAA and AATGAA mutants showed efficiency reduced to 2% (Wickens & Stefenson, 1984). Although the AAUUAA sequence occurs in 12% of all mRNAs, this mutant was also reported to have reduced 3' processing (4%) in this *Xenopus* system.

The wheat  $\alpha$  subunit mRNA contains a sequence AUCUG located upstream of the AAUAAA-like motif, which shows a 60% match with a second recognition site for polyadenylation reported in plant sequences (Table 14; Joshi, 1987). An AT-rich region located downstream of the AAUAAA-like sequence in plants is not found in the wheat mRNA (Table 14). The C nucleotide preceding the poly(A) region in the wheat sequence conforms to the observed preference of either a C or a T in known eukaryotic messengers (Birnstiel *et al.*, 1985; Joshi, 1987). The above observations indicate that efficient 3' RNA processing is ensured only when a balanced set of recognition sequences is present in the correct processing environment. This idea is supported by the notion that animal and viral poly(A) signals are improperly and inefficiently recognized in tobacco cells (Hunt *et al.*, 1987). Although the latter might not be surprising, even within the plant kingdom the mechanism by which pre-mRNA is processed is not uniform. The transcript of the gene encoding the wheat Rubisco small subunit is processed inefficiently in transgenic tobacco plants since additional sites in the wheat 3' non-coding region function as novel sites for polyadenylation (Keith & Chua, 1986).

Table 14. Nucleotide frequencies at positions surrounding the AAUAAA-like motif.

A. Nucleotide frequencies at 10 positions upstream.

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
A %	17	48	33	43	35	28	50	26	17	26
G %	17	17	17	15	20	26	2	33	24	37
C %	17	13	9	11	26	17	11	11	15	9
U %	48	22	41	30	20	28	37	30	43	28
Consensus	U	A	U	A	A	U	U	G	U	G
wheat cDNA	G	C	C	U	G	A	U	C	U	G

B. Nucleotide frequencies at 10 positions downstream.

	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
A %	46	37	48	30	46	37	39	35	33	28
G %	13	37	15	24	13	13	17	26	15	17
C %	15	7	9	11	24	11	11	20	9	22
U %	26	20	28	35	17	39	33	20	43	33
consensus	A	G	U	U	A	A	A	A	A	U
wheat cDNA	C	U	U	U	U	G	U	G	A	A

Data from Joshi (1987)



Like all sequences the wheat cDNA exhibits a particular codon usage which indicates a preference for the use of a G and C in the coding region [% (G+C) = 52], while in the non-coding region %(G+C) = 43. The strategy for choosing synonymous codons is unique to a particular species. However it is not inflexible since intraspecific variation between genes, particularly in large genomes such as those of humans, does occur. It is the average codon frequencies in all genes in the genome that characterizes the species. The nucleotide composition of pooled protein coding sequences shows an equal use of (G+C) and (A+T) in *E.coli* (48%), human (51%) and plant cells (50%; calculated from data published by Murray *et al.*, 1989). Only in yeast is a bias towards the use of A or T (60%), rather than G or C (40%), observed (Table 7). The calculated %(G+C) for protein coding sequences is about 50% even for species separated by great evolutionary distances such as *E.coli* and humans, whereas the experimentally determined %G+C for total nuclear DNA is 41% and 46% for total human and wheat germ DNA respectively (Grantham *et al.*, 1986; Marmur & Doty, 1962). These observations suggest that there is a different usage of nucleotides in coding and non-coding regions in a particular genome. This pattern is also observed on a much smaller scale i.e. in the wheat cDNA sequence where the %(G+C) for the coding region and the non-coding regions are 52% and 43% respectively (Table 8).

The codon usage of a particular species is mostly determined by the choice of nucleotides in the third position of the quartet codons. In monocot plants and *E.coli* a G or C residue is used preferentially while in yeast and dicot plants the opposite is observed (Table 10A). The low (G+C) content (44%) observed in the third position of quartet codons in the wheat cDNA coding region does not conform a typical monocot distribution. Extreme codon bias within genes in a single species has been related to the level of expression of the protein encoded by that gene. In yeast the biased codon choice towards (G+C)<sub>3</sub> in highly expressed genes such as that for phosphoglycerate kinase, enhances translation and is required to maintain mRNA stability (Hoekema *et al.*, 1987). There is also a strong positive correlation between the molar quantity of a given t-RNA in the cell and the frequency of usage of the respective codons (Ikemura, 1985). The

significance of the low  $\%(G+C)_{qIII}$  in the wheat coding region is not clear as only a high  $\%(G+C)_{qIII}$  has been reported to be correlated to a high level of expression of two genes in maize leaves i.e. the genes coding for chlorophyll a/b binding protein and the Rubisco small subunit (Table 10B; Matsuoka *et al.*, 1987). Whether the high level of expression of the Rubisco large subunit gene is also reflected in a biased codon usage remains to be determined. However, in contrast to the codon frequencies observed for nuclear genes, the  $\% (G+C)$  in the quartet codons in chloroplast genes is 23%, indicating a preference for A and T rather than for G and C (calculated from data published by Grantham *et al.*, 1986).

#### 3.2.6.4 Concluding remarks

The role of the mature  $\alpha$  subunit in the function of the Rubisco subunit binding protein oligomer is not known, but amino acid sequence comparisons with related proteins could provide more information, with special interest focused on the identified potential nucleotide binding site. Comparison of the wheat binding protein  $\alpha$  subunit sequence with that of related proteins is presented in the following chapter. The comparison of secondary and/or tertiary structures required for nucleotide binding site is likely to be more accurate in identifying an ATP binding site than do the comparison of primary structures. The secondary structure prediction of the reversed sequence in the  $\alpha$  subunit shows a speculative  $\beta$ - $\alpha$ -turn- $\beta$  conformation which might conform to the reported  $\beta$ - $\alpha$ - $\beta$  structure for the fingerprint sequence. However more direct evidence is required to identify the reversed sequence as an ATP-binding domain.

### 3.3 IDENTIFICATION AND ANALYSIS OF PROTEINS RELATED TO THE RUBISCO SUBUNIT BINDING PROTEIN ALPHA SUBUNIT OF *TRITICUM AESTIVUM*.

#### 3.3.1 INTRODUCTION

Antibodies raised against the Rubisco subunit binding protein purified from *P.sativum* leaves, and the *Triticum aestivum* cDNA nucleotide and derived amino acid sequences were used to search for related proteins in a range of organisms and databases. This study led to the conclusion that the Rubisco subunit binding protein is not confined to organisms containing Rubisco but is a member of a class of molecular chaperones found in all plastids, mitochondria and bacteria examined. This group is termed the chaperonins.

#### 3.3.2 PROTEINS RELATED IMMUNOLOGICALLY OR BY SEQUENCE TO THE RUBISCO SUBUNIT BINDING PROTEIN ALPHA SUBUNIT

##### 3.3.2.1 The *groEL* protein of *Escherichia coli*.

During the course of this work our collaborator S.M. Hemmingsen had isolated a cDNA sequence from a  $\lambda$  gt11 expression library from *Ricinus communis* endosperm using the antibodies against the Rubisco subunit binding protein purified from *P.sativum*. An initial search of the Genbank database with the castorbean nucleotide sequence revealed that it is related to the (*ams*<sup>+</sup>) gene of *E.coli*. The (*ams*<sup>+</sup>) gene, where *ams* stands for alteration of mRNA stability, had been identified for its ability to complement an *E.coli* mutant that showed a prolonged mRNA half-life (Kuwano *et al.*, 1977; Chanda *et al.*, 1985). In a telephone conversation between S.M. Hemmingsen and H.F. Kung who published the sequence of the (*ams*<sup>+</sup>) gene, H.F. Kung revealed that he had been contacted by R.A. Young who had identified sequence similarity

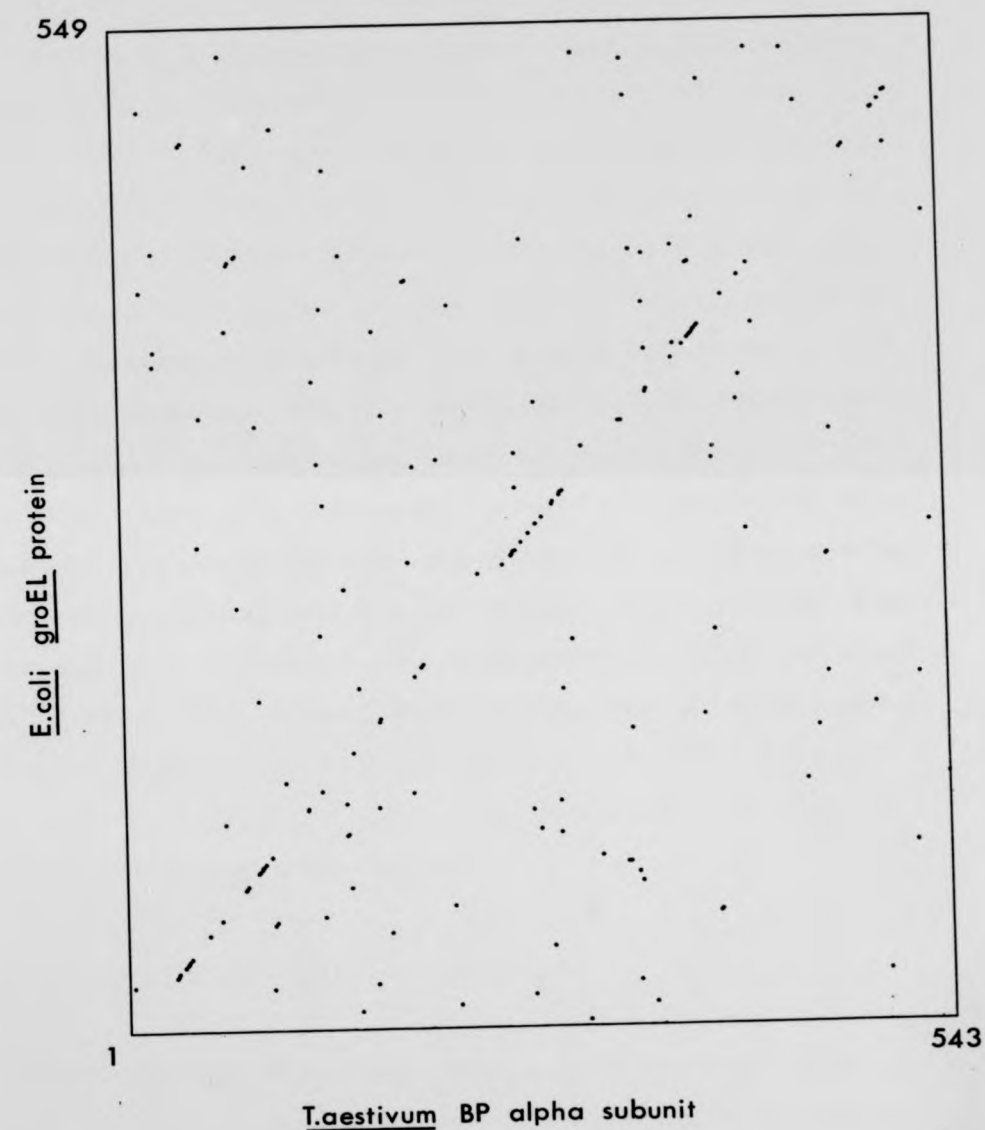
between the (*ams*<sup>+</sup>) gene and the 65 kDa common antigen of *Mycobacterium leprae*; a protein with an unknown function. However R.A. Young knew that R.W. Hendrix had identified the (*ams*<sup>+</sup>) gene as part of a sequence encoding an *E.coli* polypeptide called the *groEL* protein. Soon after this the sequence of the wheat Rubisco subunit binding protein  $\alpha$  subunit was completed and detailed amino acid comparison showed that the *E.coli groEL* protein and the wheat binding protein  $\alpha$  subunit are related as 46% of the residues are identical and many of the differences represent conserved substitutions (Fig. 29; Table 15 and Table 16).

The *groEL* protein is a product of the *groE* operon, which maps to 94 min on the *E.coli* genetic map (Georgopoulos & Eisen, 1974). The *groE* locus was discovered in studies of mutants of *E.coli* which were unable to support growth of bacteriophages  $\lambda$  and T4, and was interpreted as a host locus required for the assembly of the phage head (Takano & Kakefuda, 1972; Georgopoulos *et al.*, 1972, 1973; Sternberg, 1973). The failure of the tail of bacteriophage T5 to assemble correctly is also a result of a mutation in the *groE* locus (Zweig & Cumming, 1973). The *groE* locus is part of the *E.coli* heat shock regulon and is composed of two genes, *groEL* and *groES*. Mutations in either gene have shown to cause similar phenotypes with respect to lambda phage head morphogenesis (Tilly *et al.*, 1981) and bacterial growth (Georgopoulos & Eisen, 1974). The *groEL* and *groES* genes encode polypeptides with apparent relative molecular masses of 65 000 and 15 000 respectively, and are among the most abundant proteins in the cell (Hendrix & Tsui, 1978; Tilly *et al.*, 1981). The purified *groEL* protein is a homo-oligomer of 14 subunits and composed of two stacked rings of 7 subunits each, with a hole through the centre (Hohn *et al.*, 1979; Hendrix, 1979). The purified *groES* protein consists of 6-8 identical subunits arranged in a single ring (Chandrasekhar *et al.*, 1986).

The correct assembly of the bacteriophage lambda headstructure requires the products of at least ten phage genes (Weigle, 1966; Parkinson, 1968; Murialdo & Siminovitch, 1972a; 1972b) and the two host genes *groEL* and *groES*. Assembly intermediates have been identified among which is the 'prohead', a structure that

Figure 29. Amino acid sequence comparison of the *Triticum aestivum* Rubisco subunit binding protein  $\alpha$  subunit and the *Escherichia coli* *groEL* protein

The amino acid comparison is presented in the form of a matrix plot (Maizel & Shenk, 1981). The two amino acid sequences were compared over a moving window of 5 residues and identity was scored when 3 out of 5 amino acids were identical in any window. The numbers indicate the total number of amino acids of the sequence along a particular axis. The *groEL* amino acid sequence was taken from a paper by Hemmingsen et al. (1988)





consists of the proteins pE, pB, pB\*, pX1 and pX2 (Fig. 30); the prohead is a direct precursor to the complete head (Hohn & Hohn, 1974; Hohn *et al.*, 1974; Kaiser *et al.*, 1974; 1975). The first step in the assembly pathway of bacteriophage lambda proheads is the formation of an oligomeric structure that consists of 12 identical phage polypeptides (pB) and is known as the 'preconnector' (Fig. 30A). The correct assembly of the preconnector structure requires the 'activity' of both the host proteins, *groEL* and *groES* (Kochan & Murialdo, 1983). It is important to realize that synthesized pB polypeptides form an incorrect inactive oligomer in the absence of the *groE* proteins, and this structure can not be transformed into an active preconnector by adding the *groE* protein *in vitro* (Murialdo & Becker, 1978). Thus although the *groE* proteins participate in the assembly of the phage B polypeptides they do not form part of the final preconnector structure. A model of the preconnector is shown in Fig. 30C. It is this basic structure on which the head shell proteins, pE and pC assemble (Fig. 30B). The axial hole through the middle of the preconnector is suggested to be the entrance through which the DNA enters the prohead (Kochan *et al.*, 1984). The observed high degree of amino acid sequence identity between the Rubisco subunit binding protein  $\alpha$  subunit and the *groEL* protein (46%, Table 15), and the implied similar function of the two proteins, lead us to suspect that these two proteins may be two members of a distinct class of proteins whose function is to mediate the correct assembly of other oligomeric protein complexes.

#### 3.3.2.2 The bacterial 65 kDa common antigen

The wheat Rubisco subunit binding protein  $\alpha$  subunit also shows a high degree of amino acid similarity to a protein found in *Mycobacterium leprae* and *M.tuberculosis* (Fig. 31), with an apparent Mr of about 65 000. The percentage identical amino acids is 50%, whereas many of the differences are conserved substitutions, similar to those observed for the wheat  $\alpha$  subunit and the *E.coli groEL* protein (Table 15 and 16). The 65kDa protein is well-characterized since it elicits a strong antibody and T-cell

Figure 30. Assembly pathway of bacteriophage lambda proheads

The assembly pathway presented is based on data reported by Murialdo (1979) and Kochan & Murialdo (1983), in which pB is the monomeric phage B protein; pB<sup>\*</sup> is the oligomeric 12-mer or preconnector; pB<sup>\*</sup> is a cleaved derivative of pB; pC is a phage protein that forms a minor component of the prohead; pC' and pC'' are derived from pC; pNu3 is a phage protein that is thought to have a core or scaffolding function and is not found in the mature proheads; pE is the major coat protein; pX1 and pX2 are the result of a fusion-cleavage reaction of gpC and gpE. Panel A shows the assembly pathway of the preconnector structure; Panel B shows the subsequent steps in the formation of mature prohead. A model of the structure of the preconnector is shown in panel C (taken from Kochan *et al.*, 1984)

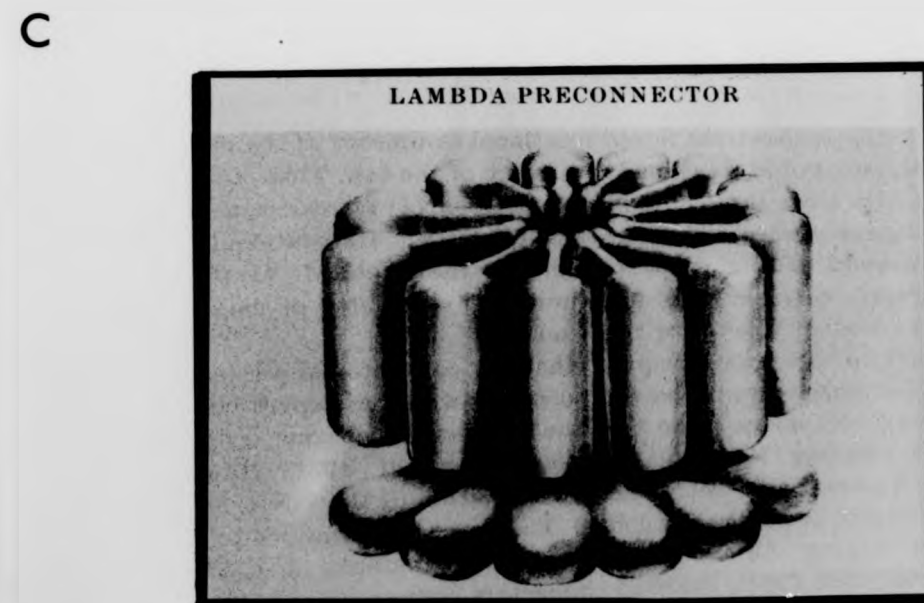
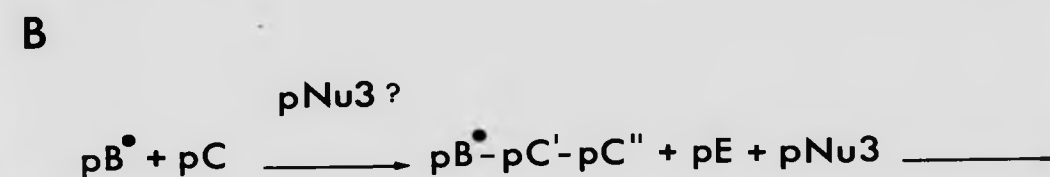
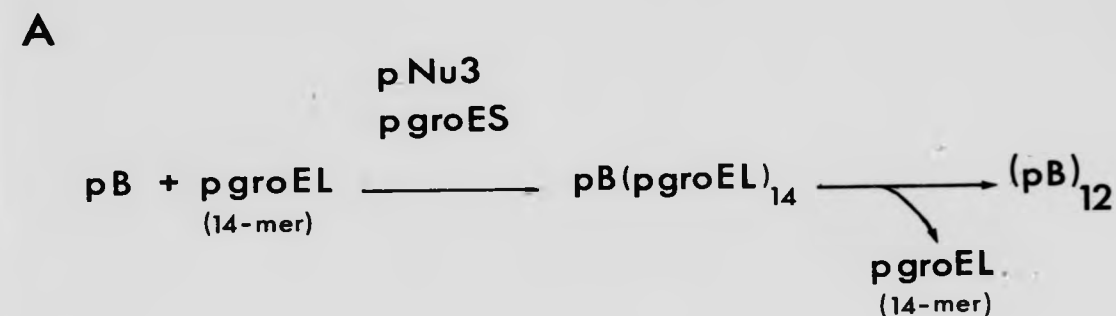
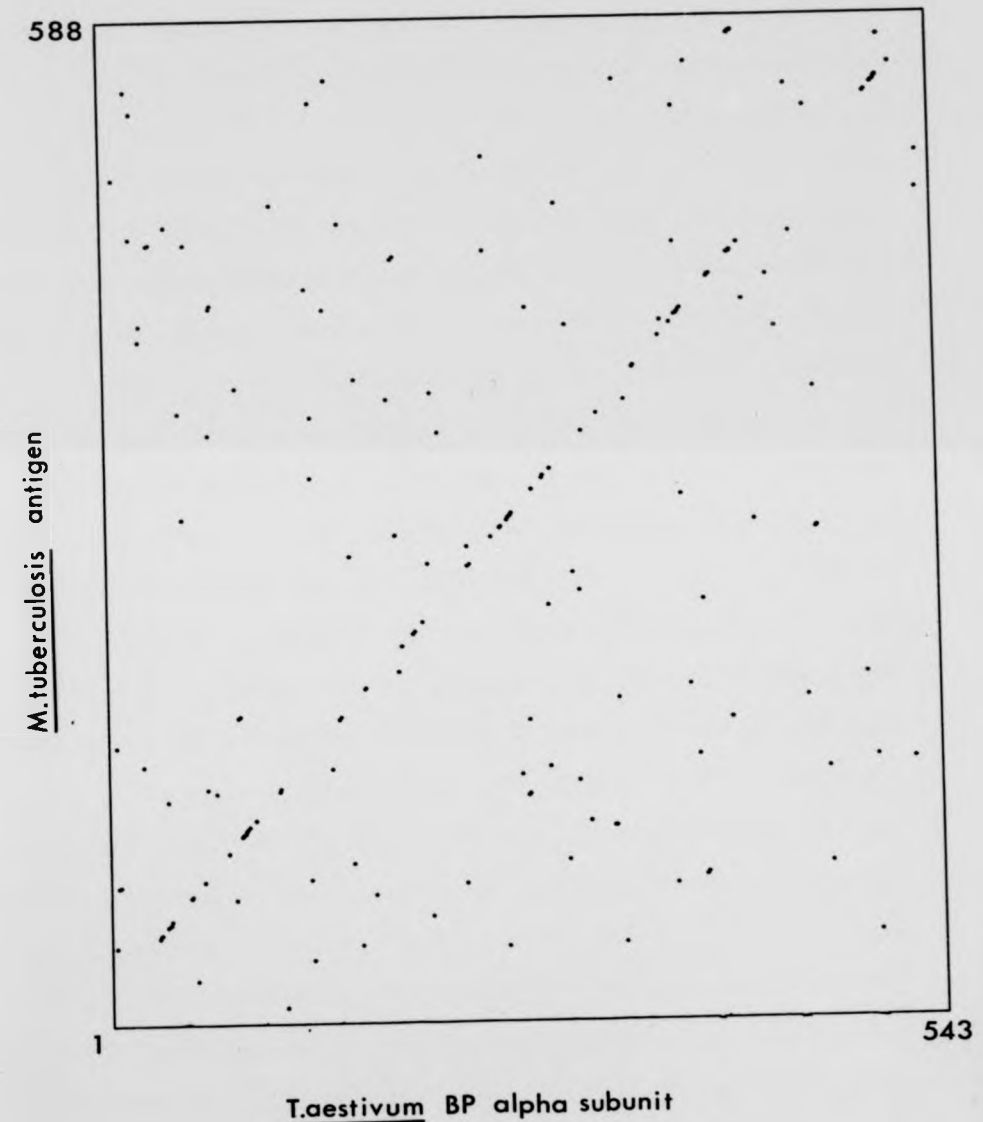


Figure 31. Amino acid sequence comparison of the *Triticum aestivum* Rubisco subunit binding protein  $\alpha$  subunit and the *Mycobacterium tuberculosis* 65 kDa antigen

The amino acid comparison is represented in the form of a matrix plot (Maizel & Shenk, 1981). A window of 5 amino acids was used and identity was scored when 3 out of the 5 residues were identical. The numbers represent the total number of amino acids in the sequence. The algorithm used was described by Needleman & Wunsch (1970). The *M. tuberculosis* amino acid sequence was taken from Shinnick (1987).





response in infected patients (Emmerich *et al.*, 1986; Gilles *et al.*, 1985). Monoclonal antibodies have been used to identify the 65 kD antigen and to determine unique epitopes in the protein from *M.tuberculosis* and *M.leprae* (Engers *et al.*, 1985). A number of monoclonal antibodies cross-react with a polypeptide with an apparent Mr of 58 000 in extracts of *E.coli*, indicating that different regions in the mycobacterial 65 kD antigen are present in the *E.coli* polypeptide (Young *et al.*, 1987). Amino acid comparison revealed that the mycobacterial protein is related to the *groEL* protein, as 59% of the residues are identical (Table 15; Young *et al.*, 1988). Immunologically related polypeptides to the 65 kD antigen and the *groEL* protein have been identified in a number of bacterial species among which is the homologue in *Pseudomonas aeruginosa* (Young *et al.*, 1987). Antibodies raised against the purified protein from *P.aeruginosa* detect related proteins in approximately 55 other bacterial species, these include gram-negative as well as gram-positive bacteria (Høiby, 1975; Kaijser, 1975; Hindersson *et al.*, 1984) and archaeobacteria (Thole, 1988). Not surprisingly, the protein is also known as the bacterial common antigen.

Antibodies raised against the purified pea Rubisco subunit binding protein, the purified *E.coli groEL* protein and the purified 65 kD common antigen from *P.aeruginosa* were used in a study to detect related proteins in a variety of organisms. The binding protein antibodies cross-reacted with polypeptides of an apparent Mr between 60 000 and 70 000 in cell extracts of *Rhodomicrobium vanielii*, *Rhodopseudomonas sphaeroides*, *Escherichia coli* and *Bacillus thuringiensis*; whereas no cross-reactivity was observed with proteins from *Solfolobus* (Fig. 32). A faint band is observed in lane 5 containing a protein extract from *Anacystis nidulans* cells. The intensity of the bands on the immunoblot is either an indication of the degree of immunological recognition, or a reflection of the different amounts of protein loaded on the gel, or both.

The antibodies to the *P.aeruginosa* common antigen, (kindly provided by J. van Embden) cross-react with the *groEL* protein from *E.coli* but not with the Rubisco subunit binding protein from pea and wheat (Fig. 33). The faint lower molecular

Figure 32. Immunologically related proteins to the *Pisum sativum* Rubisco subunit binding protein in prokaryotes

Approximately 5 ml of the different bacterial cell cultures, growing at late exponential rate, was centrifuged in a MSE 18 centrifuge at 3500 g for 15 min at 4°C. The cells were resuspended in SDS loading buffer and boiled immediately for 5 min, followed by centrifugation in an Eppendorf microfuge at 11 600 g for 5 min at room temperature. An aliquot of each supernatant, containing approximately equal amount of proteins, was loaded on a 15% polyacrylamide gel containing SDS. After electrophoresis the gel was immunoblotted using antibodies raised against the purified Rubisco subunit binding protein from *Pisum sativum* and the antibodies were detected using the enzyme-linked detection method. Lane 1, partly purified Rubisco subunit binding protein from *P. sativum* (kindly provided by J.E. Musgrove); lane 2, *Bacillus thuringiensis* HD1(BA 82B); lane 3, *Solfolobus* LM; lane 4, *Escherichia coli* CSH26; lane 5, *Anacystis nidulans* 7942; lane 6, *Rhodopseudomonas sphaeroides* ATCC 17023; lane 7, *Rhodocrobium vanielii* Rm5. The bacterial extract were kindly provided by G. MacDonald. Molecular weight markers were also separated on the polyacrylamide gel and stained with amido black after transfer to nitrocellulose. The numbers represent the positions of these markers in terms of kDa.

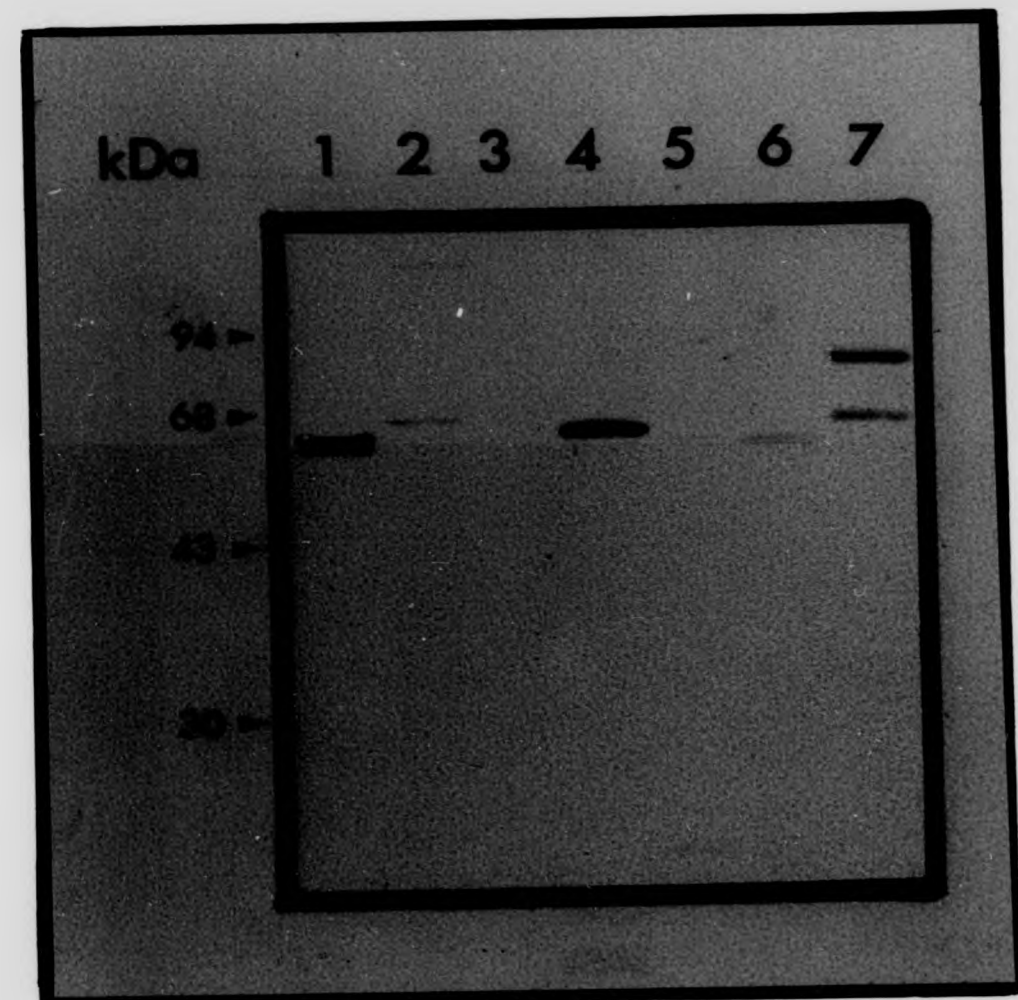
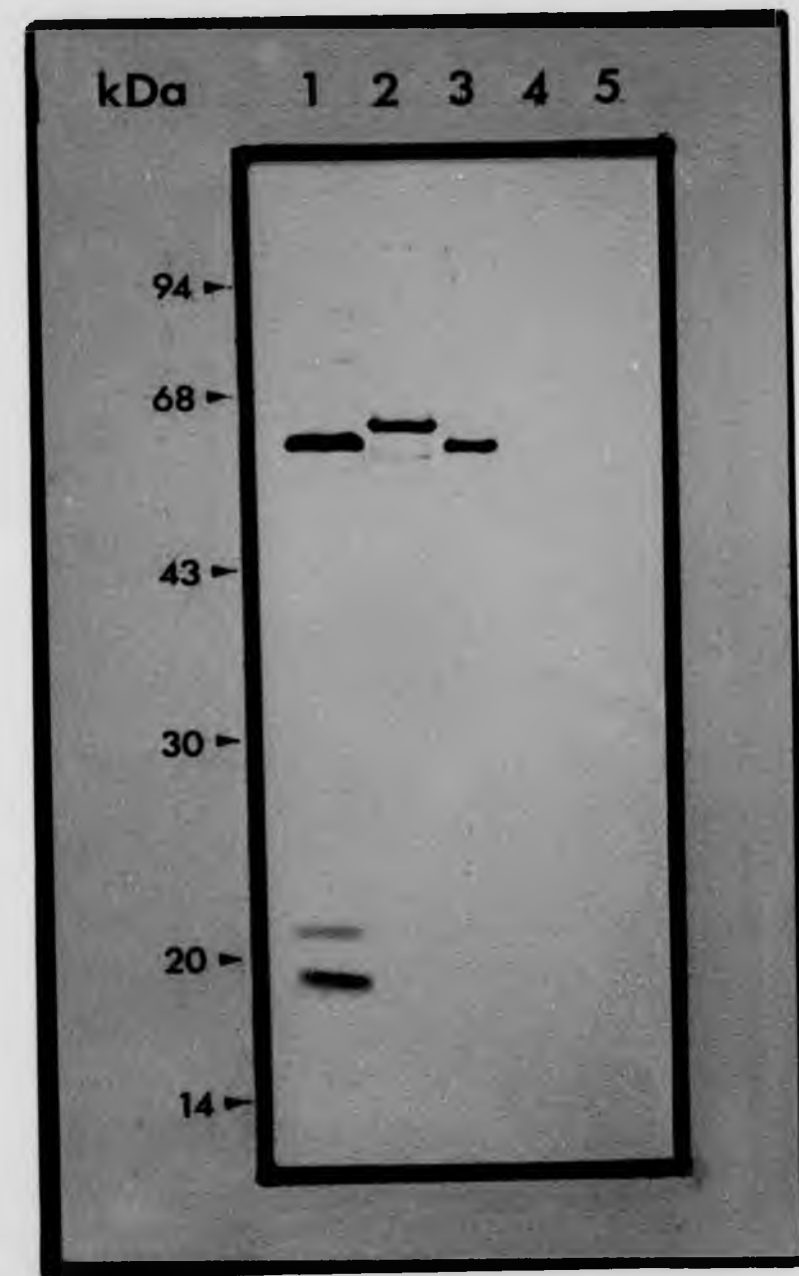


Figure 33. Immunologically related polypeptides to the *Pseudomonas aeruginosa* common antigen

Different protein samples were analyzed on a 15% polyacrylamide gel containing SDS, followed by immunoblotting. Antibodies against the *P.aeruginosa* common antigen (kindly provided by J. van Embden) were used and antibodies were detected using the enzyme-linked detection method. The cell extracts from *T.aestivum* leaves and *E.coli* (TG2) were prepared as described previously (Section 2.2.20 and Section 2.2.13.2.2). Purified *groEL* protein was a generous gift of A. Mehlert while the purified common antigen from *P.aeruginosa* was kindly provided by J. van Embden. Lane 1, total *E.coli* extract; lane 2, 65 kDa common antigen; lane 3, purified *groEL* protein; lane 4, *T.aestivum* total leaf extract and lane 5, *P.sativum* partially purified Rubisco subunit binding protein. Molecular weight markers are indicated and were determined as described in Figure 32.



weight bands observed in the control sample, i.e. the purified 65 kDa common antigen (Fig. 33, lane 2) are the result of a small amount of contaminating polypeptides in the initial purified protein preparation which was used to raise antibodies (J. van Embden, personal communication). Antibodies raised against the purified *groEL* protein (Section 2.2.13.1) recognized the common antigen from *P. aeruginosa* but neither of the binding protein subunits from wheat (Fig. 34). These results suggest that there is a higher degree of similarity between the prokaryotic polypeptides themselves than between the prokaryotic and the eukaryotic proteins. In order to confirm this observation antibodies to the pea binding protein were used to determine their cross reactivity with related proteins in prokaryotes. Cross-reactivity was observed with the *groEL* protein, the bacterial common antigen, and the homologue from *R.vaniellii* as well as with the binding protein subunits from pea and wheat (Fig. 35A). However when pre-immune serum was tested, it became clear that the rabbit had already raised antibodies against the bacterial common antigen, which cross-reacted with the other bacterial polypeptides (Fig. 35B). The lower molecular weight band running immediately below the *groEL* polypeptide, (Fig. 35A, lane 5), was often observed with preparations that had been frozen and thawed a number of times. Freshly prepared samples analyzed on a polyacrylamide gel containing SDS, followed by either Coomassie brilliant blue staining or immunoblotting using antibodies to either the pea binding protein, the 65 kD common antigen or the *groEL* protein, always showed one polypeptide.

The Rubisco subunit binding protein from pea and barley consists of two immunologically distinct polypeptides,  $\alpha$  and  $\beta$ , i.e. antibodies to the  $\alpha$  subunit do not cross-react with the  $\beta$  subunit and vice versa (Musgrove *et al.*, 1987). The possibility that one of the two polypeptides is more closely related to the prokaryotic common antigen than the other subunit was explored. The pea  $\alpha$  subunit antibodies recognized the 65 kDa common antigen and the *groEL* protein (Fig. 36A, lane 4 and 5) but failed to cross react with the common antigen from *R.vaniellii* (Fig. 36A, lane 3). The  $\beta$  subunit antibodies, on the other hand, recognized the common antigen in cell extracts

Figure 34. Immunologically related polypeptides to the *Escherichia coli groEL* protein

Protein samples were analyzed on a 15% polyacrylamide gel and immunoblotted. Antibodies raised against the purified *groEL* protein (a generous gift of A. Mehler) were used, followed by antibody detection with the aid of radiolabelled protein A (Section 2.2.14.2.2). Lane 1, purified *groEL* protein; lane 2, purified 65 kDa common antigen from *P.aeruginosa* (kindly provided by J. van Embden); lane 3, total cell extract from *T.aestivum* green leaves, prepared as described previously (Section 2.2.20). Molecular weight markers are indicated and were determined as described in Figure 32.

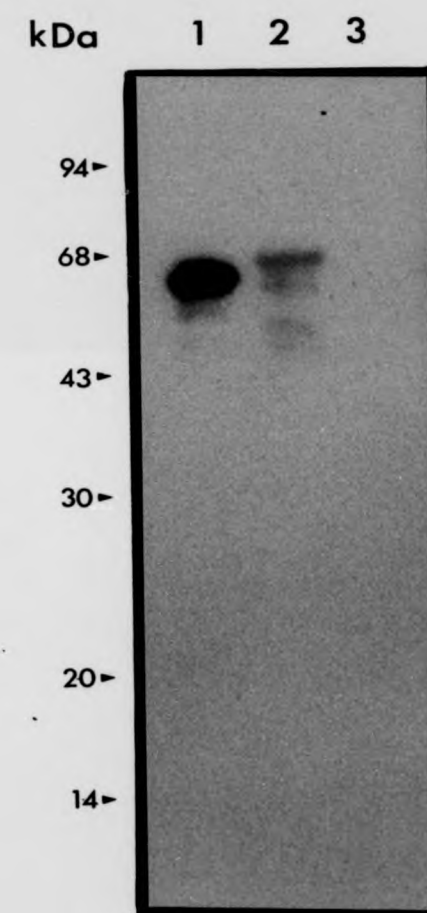




Figure 35. Cross reactivity of the antibodies to the *Pisum sativum* Rubisco subunit binding protein with polypeptides in prokaryotic cell extracts

Protein samples were analyzed on a 15% polyacrylamide gel containing SDS, followed by immunoblotting. Antibodies were detected using the enzyme-linked detection method. Samples were prepared as described in the legends of Figure 32 and Figure 33. Lane 1, partly purified Rubisco subunit binding protein from *P.sativum*; lane 2, total cell extract of *T.aestivum* leaves; lane 3, purified common antigen from *R.vaniellii* (kindly provided by G. MacDonald); lane 4, purified 65 kDa common antigen from *P.aeruginosa*, and lane 5, purified *groEL* protein from *E.coli*. Panel A shows the results of the incubation with antibodies against the purified Rubisco subunit binding protein from *P.sativum* and Panel B shows the result of incubation with the preimmune serum. Molecular weight markers are indicated and were determined as described in Figure 32.

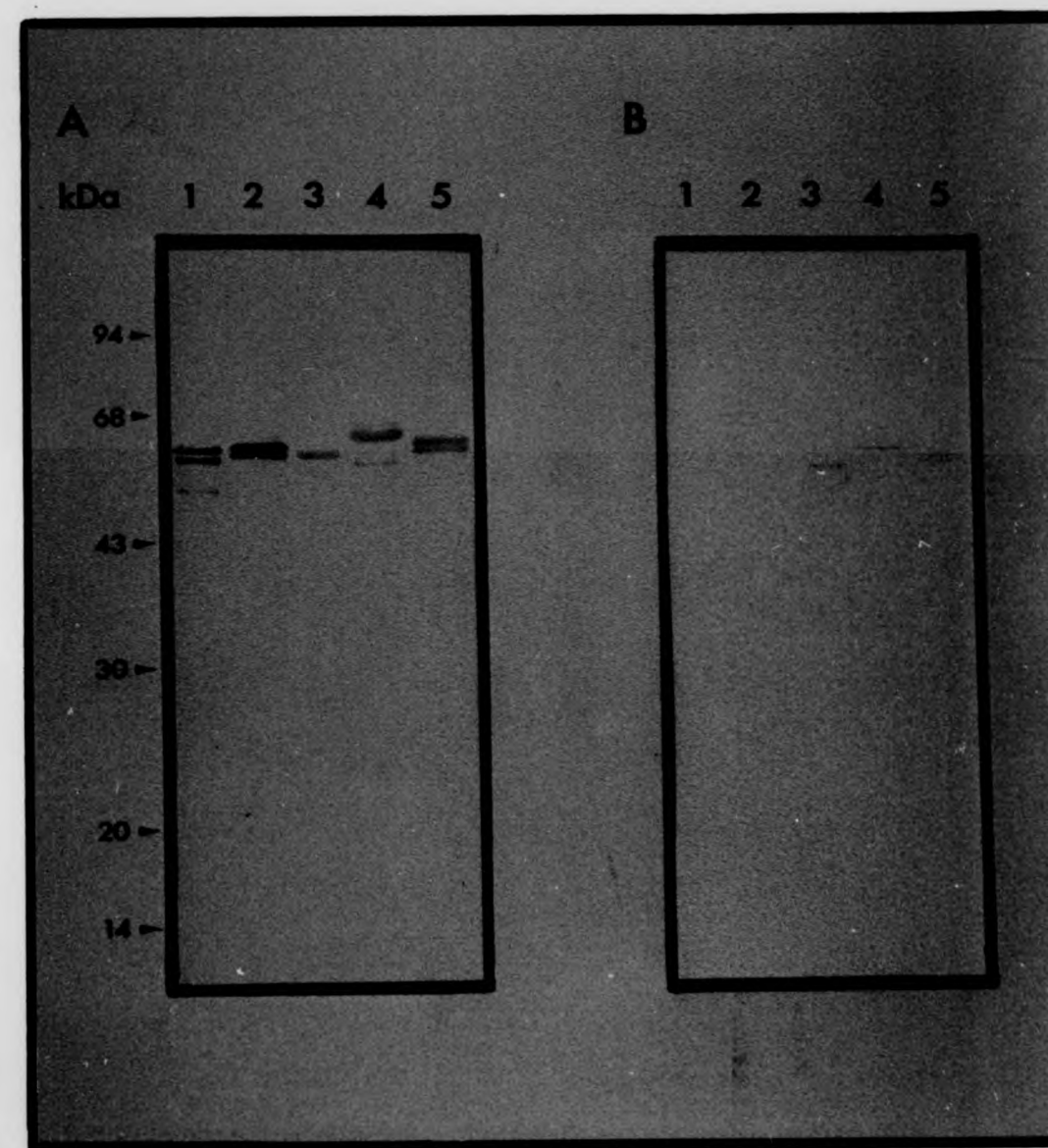
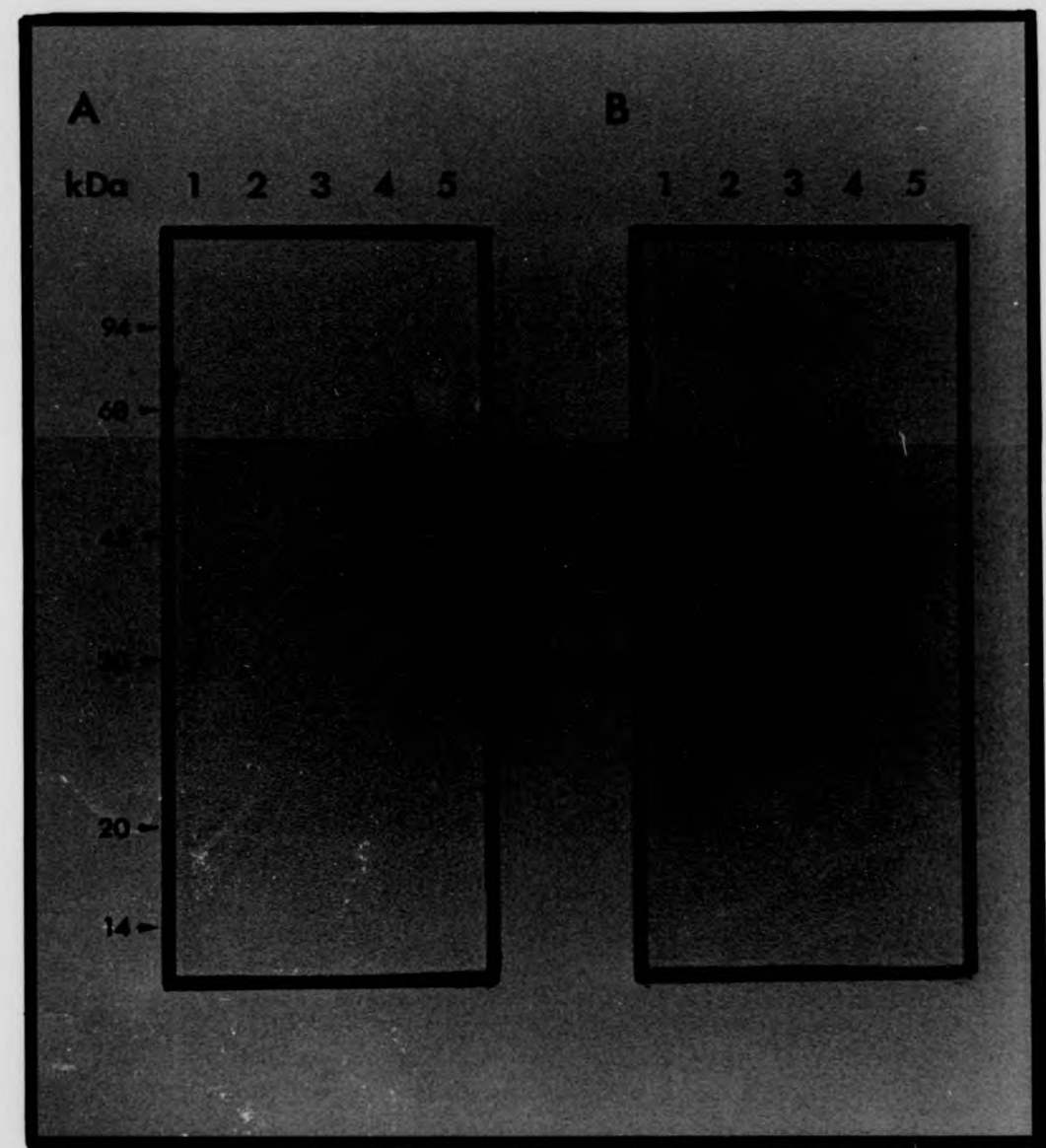


Figure 36. Cross reactivity of the antibodies to the purified  $\alpha$  and  $\beta$  subunits of the Rubisco subunit binding protein from *Pisum sativum* with prokaryotic proteins

Protein samples were prepared as described in Figure 32, Figure 33 and Figure 35, and analyzed electrophoretically on a 15 % polyacrylamide gel containing SDS, followed by immunoblotting and antibody detection using the enzyme-linked detection method. Lane 1, partly purified Rubisco subunit binding protein from *P.sativum*; lane 2, total cell extract of *T.aestivum* leaves; lane 3, purified bacterial common antigen from *R.vanelii*; lane 4, purified 65 kDa common antigen from *P. aeruginosa* and lane 5, purified *groEL* protein from *E.coli*. Panel A shows the result after incubation with antibodies against the Rubisco subunit binding protein  $\alpha$  subunit from *P.sativum* and Panel B shows the result of the incubation with antibodies against the Rubisco subunit binding protein  $\beta$  subunit from *P.sativum*. Molecular weight markers are indicated and were determined as described in the legend to Figure 32.



of *R.vanielli* as a faint band is observed (Fig. 36B, lane 3). No bands were observed in lane 4 and 5 containing the 65 kDa common antigen from *P.aeruginosa* and the *groEL* protein respectively. It was concluded that the immunologically identified polypeptides in the prokaryotes examined are probably all related.

The bacterial common antigen was also detected in a cyanobacterial protein extract of *Phormidium laminosum* (kindly provided by P. Nicholson) using antibodies to the pea binding protein and with the antibodies raised against the 65 kDa common antigen (Fig. 33, lane 1). Neither of these two antibodies recognized polypeptides in cell extracts of *Chlamydomonas reinhardtii* (Fig. 37, lane 2). Cyanobacteria normally grow photosynthetically and hence contain the Rubisco enzyme. If the assembly pathway of Rubisco in cyanobacteria is similar to that in plants, it is not surprising to detect a protein related to the Rubisco subunit binding protein (Fig. 32, lane 5). It was noticed that all prokaryotic extracts examined contain one cross-reacting polypeptide only, in contrast to the two polypeptides observed in all plant extracts. Definitive evidence that cyanobacteria contain a *groEL*-like protein has come from the discovery that URF3 in the paper of Cozens and Walker (1987), describing the ATP synthase operon in *Synechococcus*, contains a portion of the *groEL*-like protein sequence (Cookson *et al.*, 1989). Further investigation reveals that URF3 encodes the first 300 amino acids of the cyanobacterial *groEL* protein which shows 54% amino acid identity when compared with the corresponding region of the Rubisco subunit binding protein  $\alpha$  subunit from wheat (Fig. 38; Table 15 and 16).

### 3.3.2.3 A *groEL*-like protein in mitochondria of *Tetrahymena thermophila* and *Saccharomyces cerevisiae*

An abundant protein with an apparent subunit Mr of 58 000 had been purified from *Tetrahymena thermophila* and *Saccharomyces cerevisiae* and was shown to be related to the *groEL* protein both immunologically and structurally (McMullin & Hallberg, 1987; 1988). The nuclear encoded *groEL*-like protein is exclusively located



Figure 37. Immunologically related proteins in the cyanobacterium *Phormidium laminosum*

Aliquots of 30  $\mu$ l of cell extracts of *Phormidium laminosum* and *Chlamydomonas reinhardtii* CW15 were loaded on a 15% polyacrylamide gel containing SDS. After electrophoresis the polypeptides were immunoblotted and antibodies were detected using the enzyme-linked detection method. The *P.laminosum* cells were kindly grown by P. Nicholson (Cambridge University). At stationary phase an aliquot of 5 ml was centrifuged in a MSE 21 centrifuge at 3500 g for 15 min at 4°C, and the cells in the pellet resuspended in 500  $\mu$ l SDS loading buffer, and sonicated, twice for 30 s each at room temperature, then boiled and recentrifuged. Lane 1, *P.laminosum* cell extract; lane 2, *C. reinhardtii* cell extract. Panel A shows the result after incubation with the *P.sativum* Rubisco subunit binding protein antibodies and in Panel B the result is shown after incubation with the antibodies against the *P.aeruginosa* common antigen. Molecular weight markers are indicated and were determined as described in the legend to Figure 32.

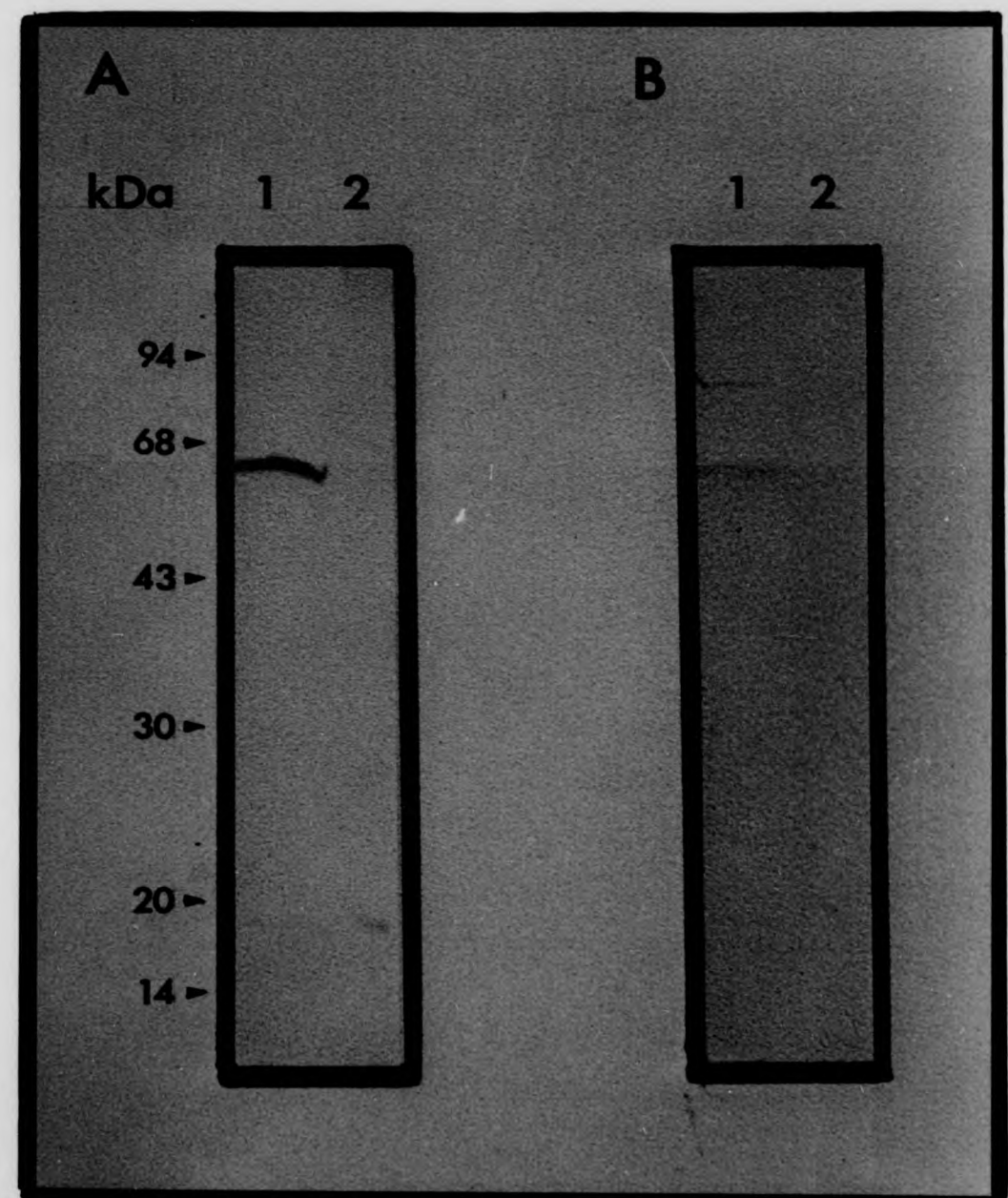
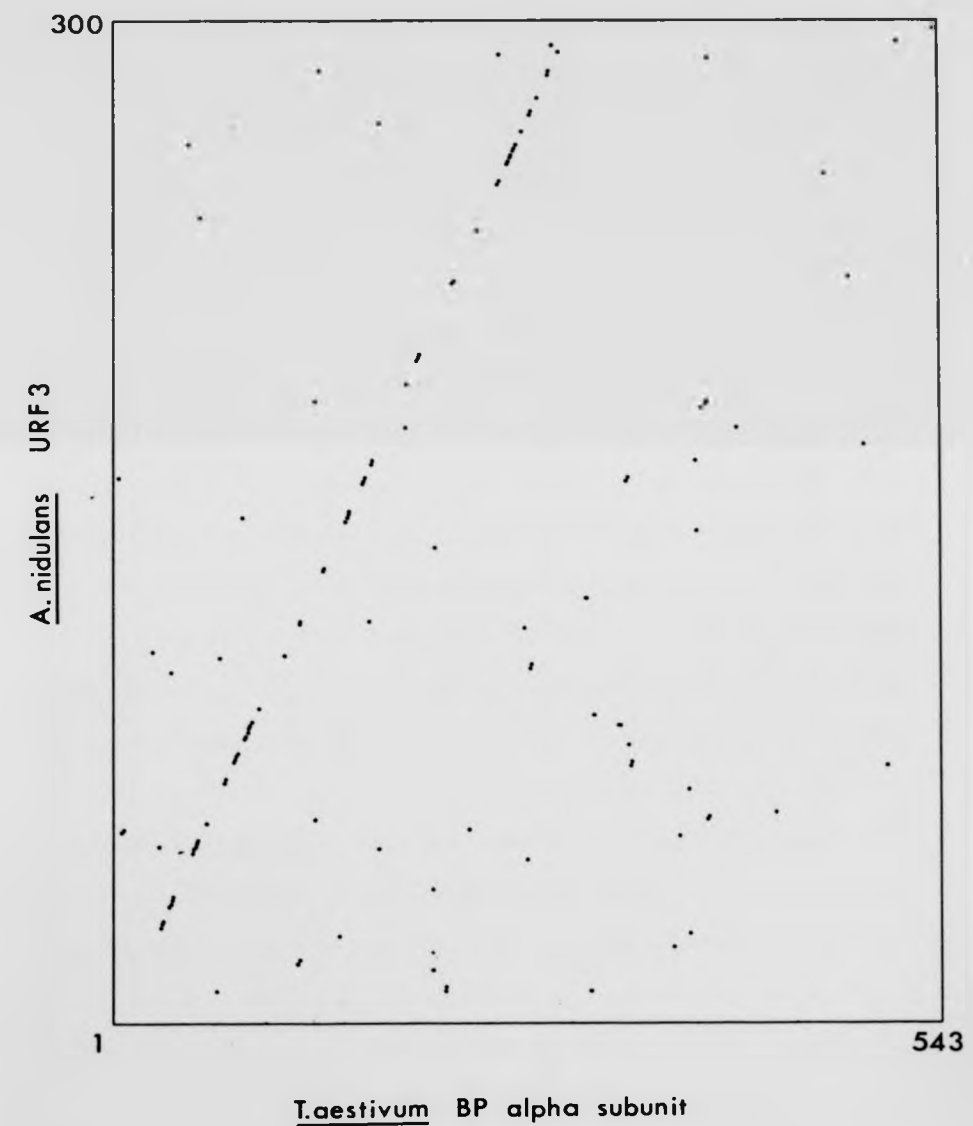


Figure 38. Amino acid sequence comparison of the *Triticum aestivum* Rubisco subunit binding protein  $\alpha$  subunit and the *Anacystis nidulans* groEL-like protein

Amino acid sequence comparison is presented in the form of a matrix plot (Maizel & shenk, 1981). A window of 5 residues was used and identity scored when 3 out of 5 amino acids were identical in any window. The numbers indicate the total number of amino acid residues in a sequence, URF = unidentified reading frame. The Needleman and Wunsch (1970) algorithm was used. The *A. nidulans* amino acid sequence was taken from Cozens and Walker (1987)



in the mitochondria where its suggested function is to assist in the refolding of newly-imported proteins. Its synthesis is induced upon heatshock and antibodies to the *T.thermophila* *groEL*-like protein cross react with a polypeptide of apparent Mr of 58 000 - 64 000 in extracts of mitochondria from *Saccharomyces cerevisiae*, *Xenopus laevis*, *Zea mays* and *Homo sapiens*, suggesting that it occurs in all mitochondria. The antibodies to the *T.thermophila* protein (kindly provided by R.L. Hallberg) were used to determine the cross-reactivity with the chloroplast Rubisco subunit binding protein. For this purpose intact chloroplasts were isolated from green leaves of *P.sativum* (Section 2.2.18.1). The stroma was separated from the membranes by centrifugation of the lysed chloroplasts. Both fractions were analyzed by polyacrylamide gel electrophoresis under denaturing conditions followed by immunoblotting.

It was found that the antibodies to the mitochondrial *T.thermophila* *groEL*-like protein cross-react with proteins running with the mobility of the chloroplast binding protein from *P.sativum* (Fig. 39). A certain amount of binding protein was detected in the membrane fraction (Fig. 39, lane 3) due to the fact that the membranes had not been washed. Since the amount of binding protein in this fraction is small, it can be seen that the antibodies to the mitochondrial protein recognized both the  $\alpha$  and  $\beta$  subunits of the chloroplast protein (Fig. 39, lane 3). It was concluded that both the  $\alpha$  and the  $\beta$  subunits are immunologically related to the mitochondrial *groEL*-like protein. The antibodies to the mitochondrial *groEL*-like protein also recognized one polypeptide with an apparent Mr of about 60 000 in cell extracts of *Triticum aestivum*, *Escherichia coli*, *Chlamydomonas reinhardtii*, *Phormidium laminosum* and *Prochlorothrix hollandica* (Fig. 36). (The latter three cell extracts were kindly provided by C.J. Howe). *Prochlorothrix hollandica* is a unicellular filamentous prokaryote and is considered to be the closest related organism to chloroplasts. In the wheat leaf extract (Fig. 40, lane 1) only one radiolabelled band is observed, which suggests that either the *T.thermophila* hsp58 antibodies do not recognize both the  $\alpha$  and  $\beta$  subunits of the Rubisco subunit binding protein from wheat or that the protein concentration on the gel

Figure 39. Immunological relationship between a mitochondrial *groEL*-like protein from *Tetrahymena thermophila* and the Rubisco subunit binding subunit from *Pisum sativum*

Chloroplasts from *P.sativum* were isolated as described in Section 2.2.18.1. Chloroplasts were lysed in 10 mM Tris-HCl pH 7.8 containing 5 mM MgCl<sub>2</sub>, and the membranes separated from the soluble proteins by centrifugation in a Beckman TL 100 centrifuge at 90 000 g for 10 min at 4°C. The supernatant fraction was diluted with an equal volume of SDS loading buffer and boiled immediately. The pellet was drained and resuspended in 10 mM Tris-HCl pH 7.8. An equal volume of SDS loading buffer was added and the suspension boiled immediately for 5 min. The stromal and membrane samples were centrifuged in an Eppendorf microfuge at 11 600 g for 5 min at room temperature, prior to loading on a 15% polyacrylamide gel containing SDS. After electrophoresis the polypeptides were transferred to nitrocellulose and incubated with antibodies raised against the purified mitochondrial *groEL*-like protein from *T.thermophila* (kindly provided by R.L. Hallberg). Antibodies were detected using radiolabelled protein A and the autoradiograph is shown. Lane 1, total chloroplast fraction; lane 2, stromal fraction; lane 3, membrane fraction. The arrows indicate the Rubisco subunit binding protein  $\alpha$  and  $\beta$  subunits. The molecular weight markers shown were determined as described in Figure 32.

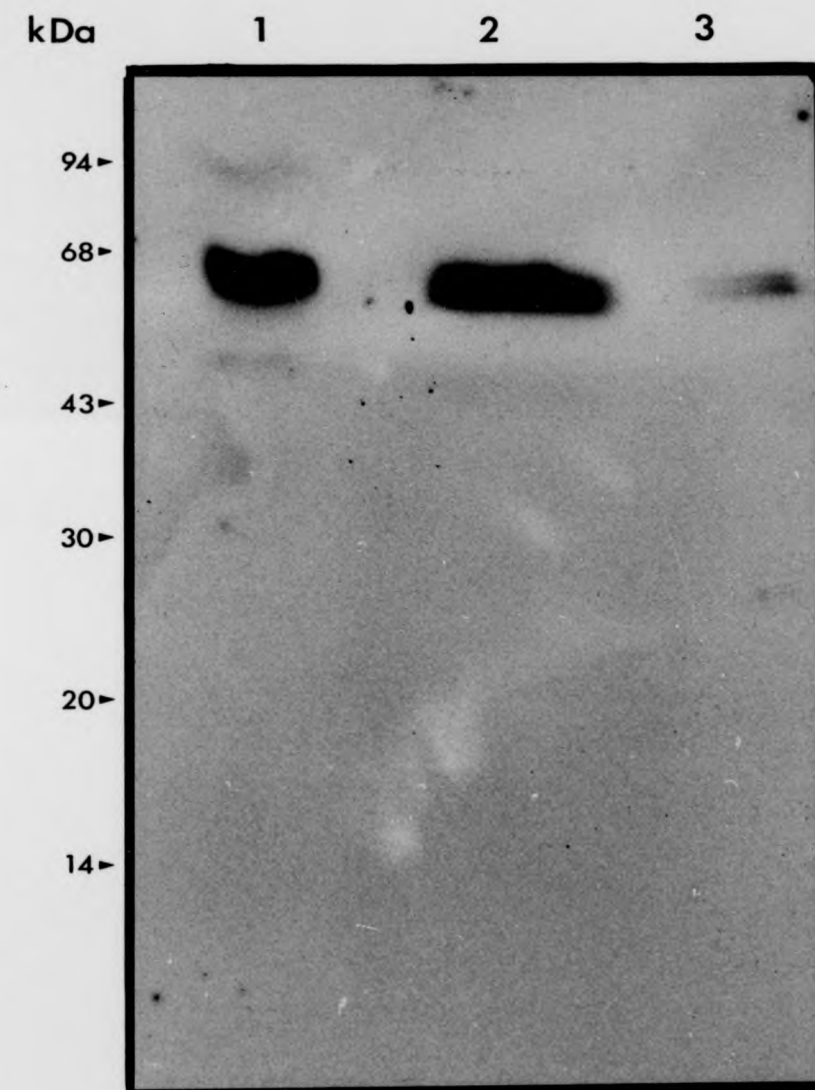
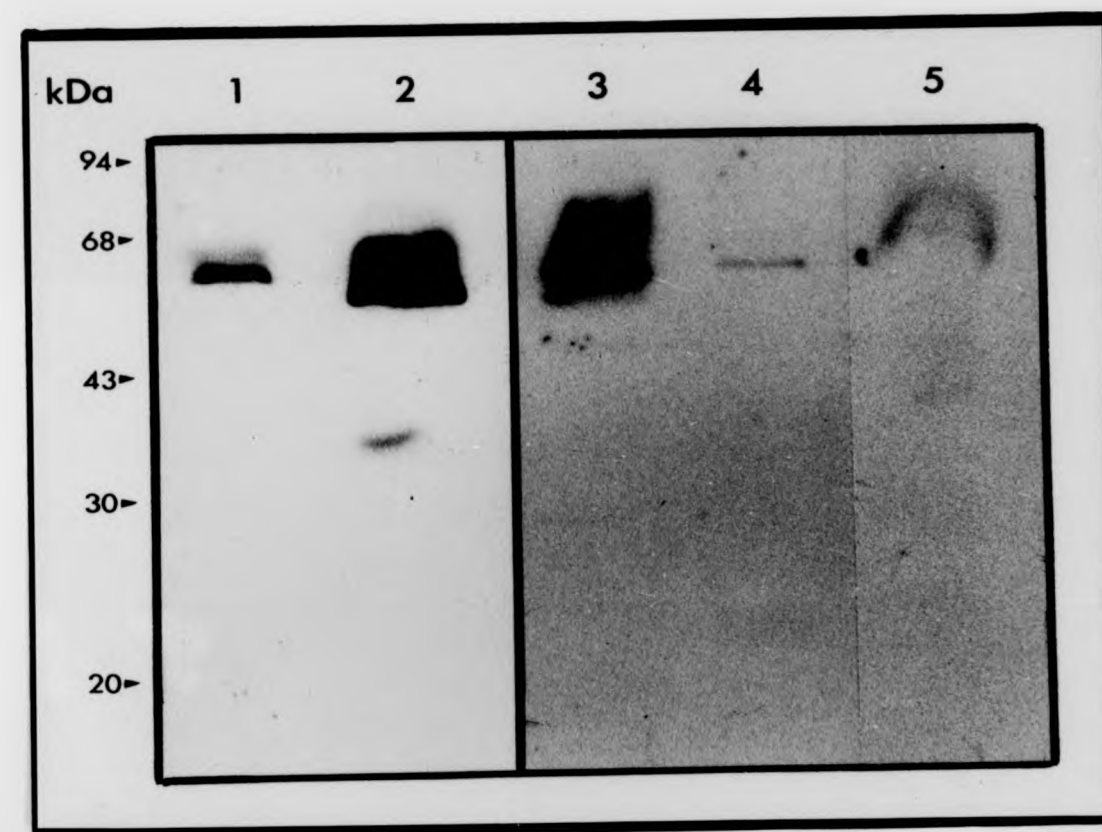


Figure 40. Immunologically related proteins to the *groEL*-like protein from *Tetrahymena thermophila*

The autoradiograph of an immunoblot incubated with antibodies against the *T.thermophila* *groEL*-like protein is shown. Protein samples were prepared as describes in Section 2.2.20, the *T.aestivum* leaf extract was prepared as described in Section 2.2.13.2.2 and the *E.coli* extract as described in Figure 37. Lane 1, *T.aestivum*; lane 2, *E.coli* (TG2); lane 3, *C.reinhardtii*; lane 4, *P.laminosum* and lane 5, *P.hollandica*. The *C.reinhardtii*, *P.laminosum* and *P.hollandica* cell extracts were kindly provided by C.J. Howe (Cambridge University). The indicated molecular weight markers were determined as described in Figure 32.





and the radiolabelled antibody detection method used do not allow the resolution of the two polypeptides. The latter interpretation is more likely to be correct, as the cross reactivity of both the Rubisco subunit binding protein subunits from stromal extracts of *Pisum sativum* chloroplasts is observed only at a low protein concentration (Fig. 39, lane 3).

It was concluded from this study that immunologically related proteins to the Rubisco subunit binding protein are present in all organisms examined. Comparison of the amino acid sequences available for some of these immunologically related proteins shows that about 50% of their amino acids are identical (Table 15). Two of these proteins, the *groEL* protein from *E.coli* and the Rubisco subunit binding protein present in plant chloroplasts, are associated with the post-translational assembly of two structurally distinct oligomeric protein complexes, i.e. the preconnector, a structure in bacteriophage  $\lambda$  proheads, and the Rubisco enzyme respectively. It was concluded that the Rubisco subunit binding protein, the *groEL* protein and the bacterial common antigen are members of a class of proteins whose role is to assist other polypeptides to maintain or assume conformations that permit their correct assembly into oligomeric structures. This conclusion has been published in a joint paper with co-workers in other laboratories (Hemmingsen *et al.*, 1988; copy included at the back of this thesis). This new class of proteins we have named 'the chaperonins'. The chaperonins are defined as a class of sequence-related molecular chaperones (see also Section 1.3.5) found in plastids, mitochondria and prokaryotes. In the remainder of this thesis this term will be used.

### 3.3.3 IDENTIFICATION OF THE CHAPERONIN IN MITOCHONDRIA OF *SOLANUM TUBEROSUM* AND *PISUM SATIVUM*

The *Tetrahymena thermophila* mitochondrial chaperonin antibodies were used to detect the chaperonin in mitochondrial fractions prepared from *Pisum sativum* leaves and *Solanum tuberosum* tubers (potato). The mitochondrial fractions (kindly provided

by A. Liddell and C.J. Leaver) were analyzed on polyacrylamide gels containing SDS followed by immunoblotting. The total potato mitochondrial fraction contains an immunologically related protein with an apparent Mr of about 60 000 referred to as chaperonin hereafter, which appears to be sensitive to degradation (Fig. 41A, lane 1). About 90% of the chaperonin was calculated to be associated with the mitochondrial membranes, based on the total amount of protein loaded on the gel in relation to the intensity of the immunological signals observed on the autoradiogram. When the membranes were washed with sodium carbonate, virtually all of the protein was released (Fig. 41A, lanes 4 and 5). Treatment with sodium carbonate will produce open sheets of membranes and effectively release peripheral membrane proteins such as the  $\alpha$  and  $\beta$  subunits of the ATPase (Fujiki *et al.*, 1982; Fig. 41B, lane 5). However in a repeat analysis of mitochondrial fractions a larger proportion of the chaperonin was detected in the matrix fraction (Fig. 41C, lane 2), but as found previously the membrane-associated protein was completely released upon treatment with sodium-carbonate (Fig. 41C, lanes 4 and 5).

Analysis of mitochondrial fractions from *P. sativum* shows that most of the chaperonin protein is associated with the membrane fraction and released upon treatment with sodium carbonate (Fig. 42). The distributions is thus similar to that observed for the mitochondrial chaperonin in potato. These observations suggest that in pea and potato the mitochondrial chaperonin is loosely associated with the membranes and the conditions used during the isolation of the mitochondrial fraction are critical for this distribution.

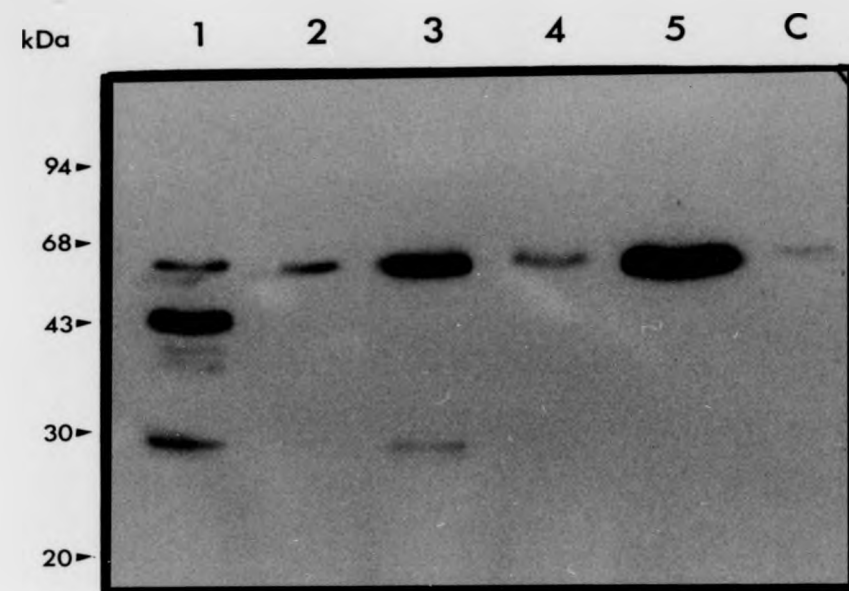
The membrane associated properties of the mitochondrial chaperonin were further examined in more detail. Sodium carbonate treatment is considered to be both non-denaturing and effective in disaggregating proteins, but lacks the charge neutrality of some detergents, thus affecting the charge properties of the solubilized proteins. This property makes it an inappropriate reagent if subsequent analysis of the proteins under non-denaturing conditions is required. The reagent 3-[(3-cholamido propyl)dimethylammonio]-1-propanesulfonate (CHAPS) seems ideal for this purpose

**Figure 41. Identification of the chaperonin in isolated mitochondria from *Solarum tuberosum***

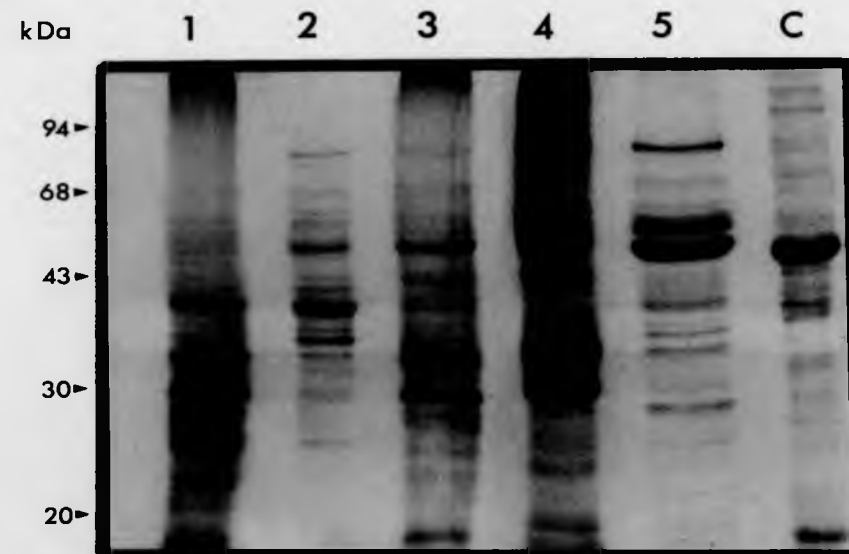
Mitochondria were isolated from potato tubers by A. Liddell using the method described by Leaver et al. (1983). The isolated mitochondria were sonicated in 10 mM MOPS pH 7.2 containing 1mM PMSF and the membranes were separated by centrifugation in a Beckman TL 100 centrifuge at 540 960 g for 30 min at 20°C. The membrane-containing pellet was resuspended in 100 mM Na<sub>2</sub>CO<sub>3</sub> pH 11,5 and left on ice for 30 min, and the fraction recentrifuged. Aliquots of the different fractions were boiled immediately in SDS loading buffer and analyzed by electrophoresis through a 15% polyacrylamide gel containing SDS, followed by immunoblotting and incubation with antibodies against the *T.thermophila* mitochondrial chaperonin protein. Antibodies were detected using radiolabelled protein A. Lane 1, total mitochondrial fraction; lane 2, mitochondrial matrix fraction; lane 3, membrane fraction; lane 4, membrane fraction after Na<sub>2</sub>CO<sub>3</sub>-wash and lane 5, supernatant fraction after Na<sub>2</sub>CO<sub>3</sub> -wash. The control is a *T.aestivum* leaf extract (Lane C). Panel A shows the autoradiograph while Panel B shows the Coomassie brilliant blue stained polyacrylamide gel after immunoblotting. Panel C shows the autoradiograph of a second similar experiment. The molecular weight markers are indicated and were determined as described in the legend to Figure 32.



**A**



**B**



**C**

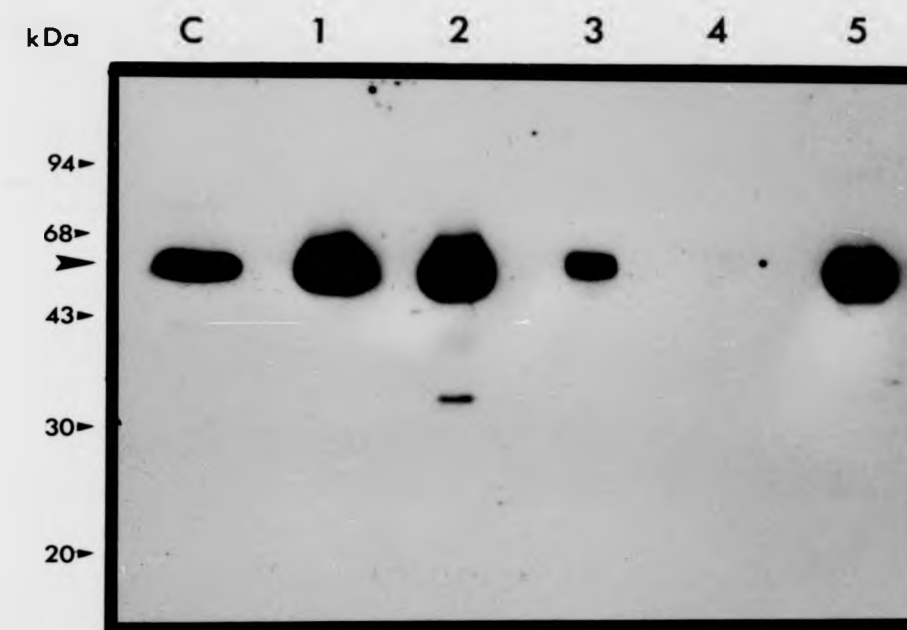
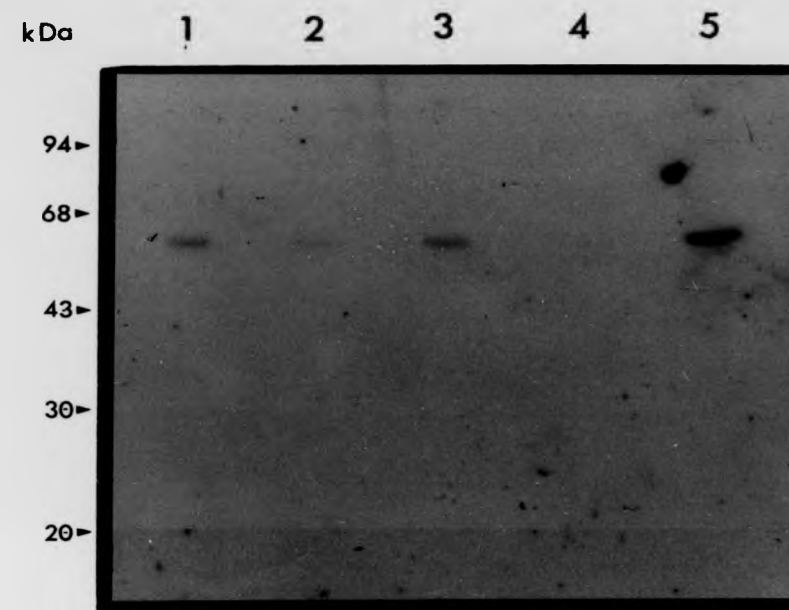


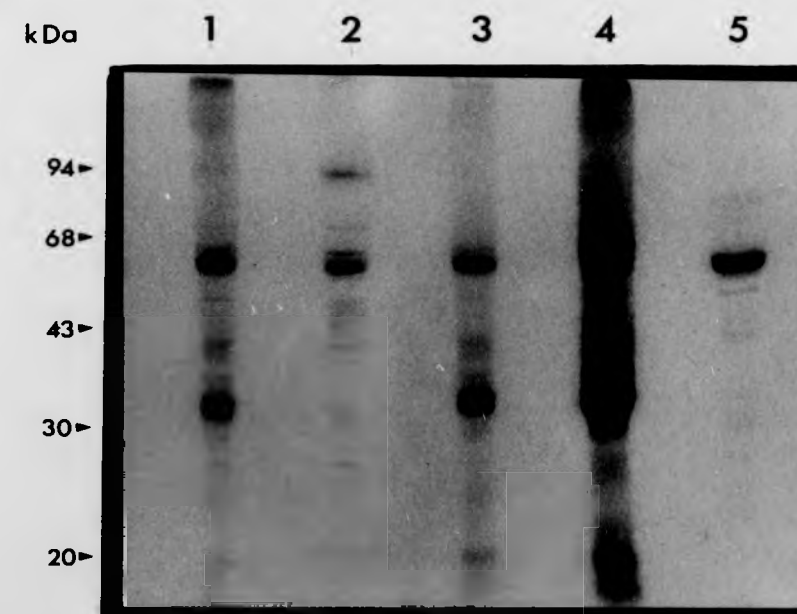
Figure 42. Identification of the chaperonin in isolated mitochondria from *Pisum sativum*

Mitochondrial fractions were prepared from pea leaves as described in Figure 41 by A. Liddell. Lane 1, total mitochondrial fraction; lane 2, mitochondrial matrix fraction; lane 3, mitochondrial membrane fraction after  $\text{Na}_2\text{CO}_3$ -wash and lane 5, supernatant fraction after the  $\text{Na}_2\text{CO}_3$ -wash. Panel A shows the autoradiograph of the immunoblotted 15% polyacrylamide gel containing SDS, after incubation with antibodies raised against the *T.thermophila* mitochondrial chaperonin. Panel B shows the Coomassie brilliant blue stained gel after immunoblotting. Molecular weight markers are indicated and were determined as described in Figure 32.

**A**



**B**



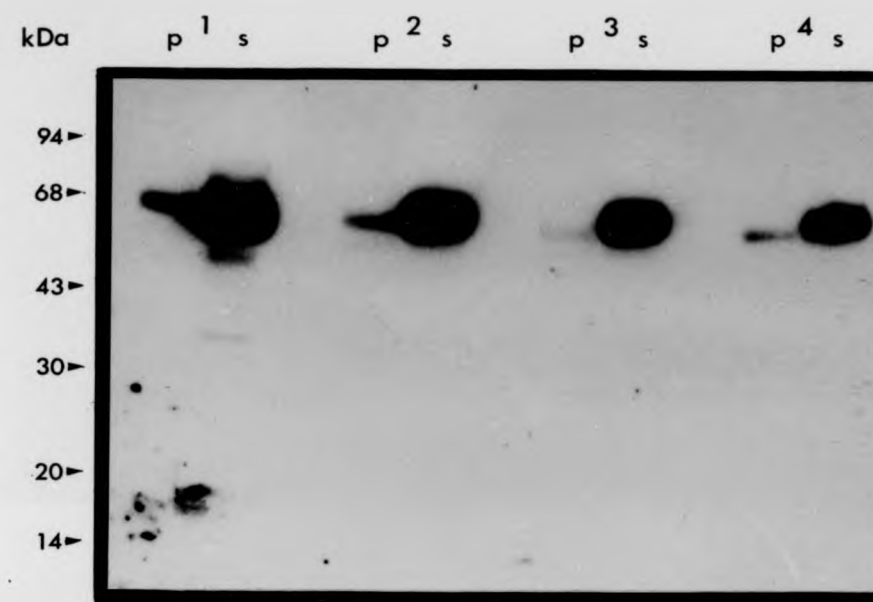
since it is electrically neutral and effective at solubilizing membrane-associated proteins in a non-denatured state (Hjelmeland, 1980; C.J. Leaver, personal communication). When potato mitochondrial membranes were washed with CHAPS, at a final concentration of 5 mM, about 80% of the chaperonin was released (Fig. 43A, lane 1). At a concentration of 40 mM CHAPS virtually 100% of the chaperonin was solubilized (Fig. 43A, lane 3). Note that at the latter concentration the amount of solubilized protein loaded on the polyacrylamide gel is much smaller than the amount of membrane proteins (Fig. 43B, lane, cf. p & s). Similar results were obtained with the mitochondrial membranes from pea (Fig. 44). It was noticed that the pea mitochondrial chaperonin is also sensitive to degradation (Fig. 44A, lane 1). At a concentration of 40 mM CHAPS and less the protein was solubilized while at a concentration of 70 mM CHAPS it was found associated with the membranes (Fig. 44A, lane 3 and lane 4, respectively). It is not obvious why this is, but the amount of solubilized protein loaded (70 mM CHAPS) was very low (Fig. 44, lane 4) and in addition almost all of the  $\alpha$  and  $\beta$  subunits of the ATPase, both extrinsic polypeptides, were detectable in the membrane fraction (Fig. 40B, lane 4). It was concluded that the best conditions to purify the mitochondrial chaperonin are those in which it is initially associated with the membrane fraction and subsequently released at a concentration of 40 mM CHAPS. All the mitochondrial fractions analyzed in Fig. 43 and 44 were kindly provided by A. Liddell and C.J. Leaver (University of Edinburgh).

Recently the mitochondrial chaperonin from *Zea mays* etiolated seedlings has been purified and was shown to be structurally related to the *groEL* protein from *E.coli* (Prasad & Hallberg, 1989). In this latter work it is stated that the antibody raised against the common antigen from *T.thermophila* was used to isolate the maize mitochondrial chaperonin because it does not recognize the homologues in chloroplast and etioplast extracts from respectively light- and dark-grown pea and maize seedlings. This observation is in contrast to our results using the same antibody (kindly provided by R.L. Hallberg) which does detect the chaperonin  $\alpha$  and  $\beta$  subunits in isolated pea chloroplasts (Fig. 39). A possible explanation is that the antibody detection method

Figure 43. Release of extrinsic proteins from mitochondrial membranes isolated from *Solarum tuberosum*

The mitochondrial membrane fraction was isolated as described in the legend to Figure 41. The pelleted membranes were resuspended in CHAPS at a final concentration of 5mM (lane 1), 15 mM (lane 2), 40 mM (lane 3), 70 mM (lane 4), and the extrinsic proteins separated by centrifugation. Both the pellet membrane fraction (p) and the solubilized protein fraction (s) were boiled in SDS loading buffer and analyzed on a 15% polyacrylamide gel containing SDS, followed by immunoblotting and incubation with antibodies against the *T.thermophila* mitochondrial chaperonin. The different mitochondrial samples were kindly provided by A. Liddell and C.J. Leaver (University of Edinburgh). Antibodies were detected using radiolabelled protein A and the autoradiograph is shown in Panel A, whereas the Coomassie brilliant blue stained gel after immunoblotting is presented in Panel B. Molecular weight markers are indicated and were determined as described in the legend to Figure 32

A



B

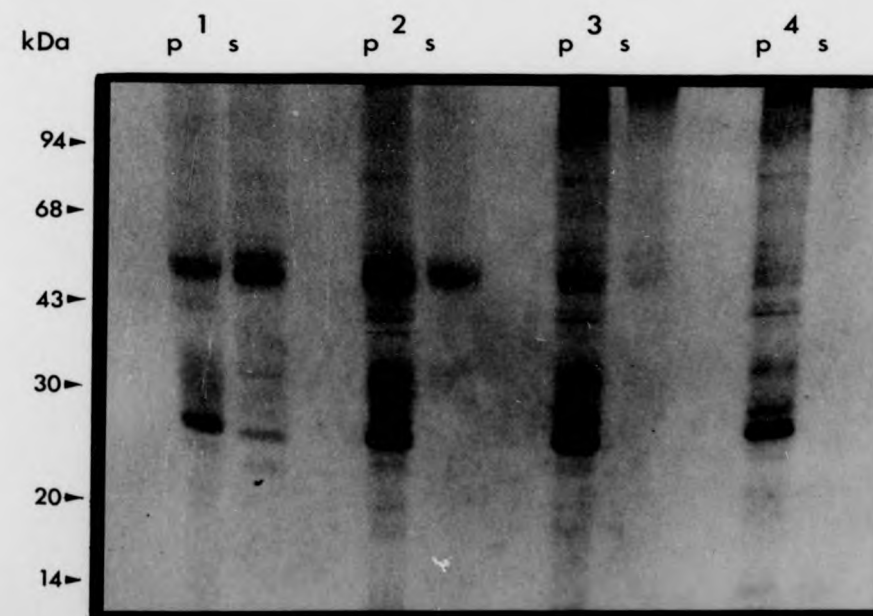
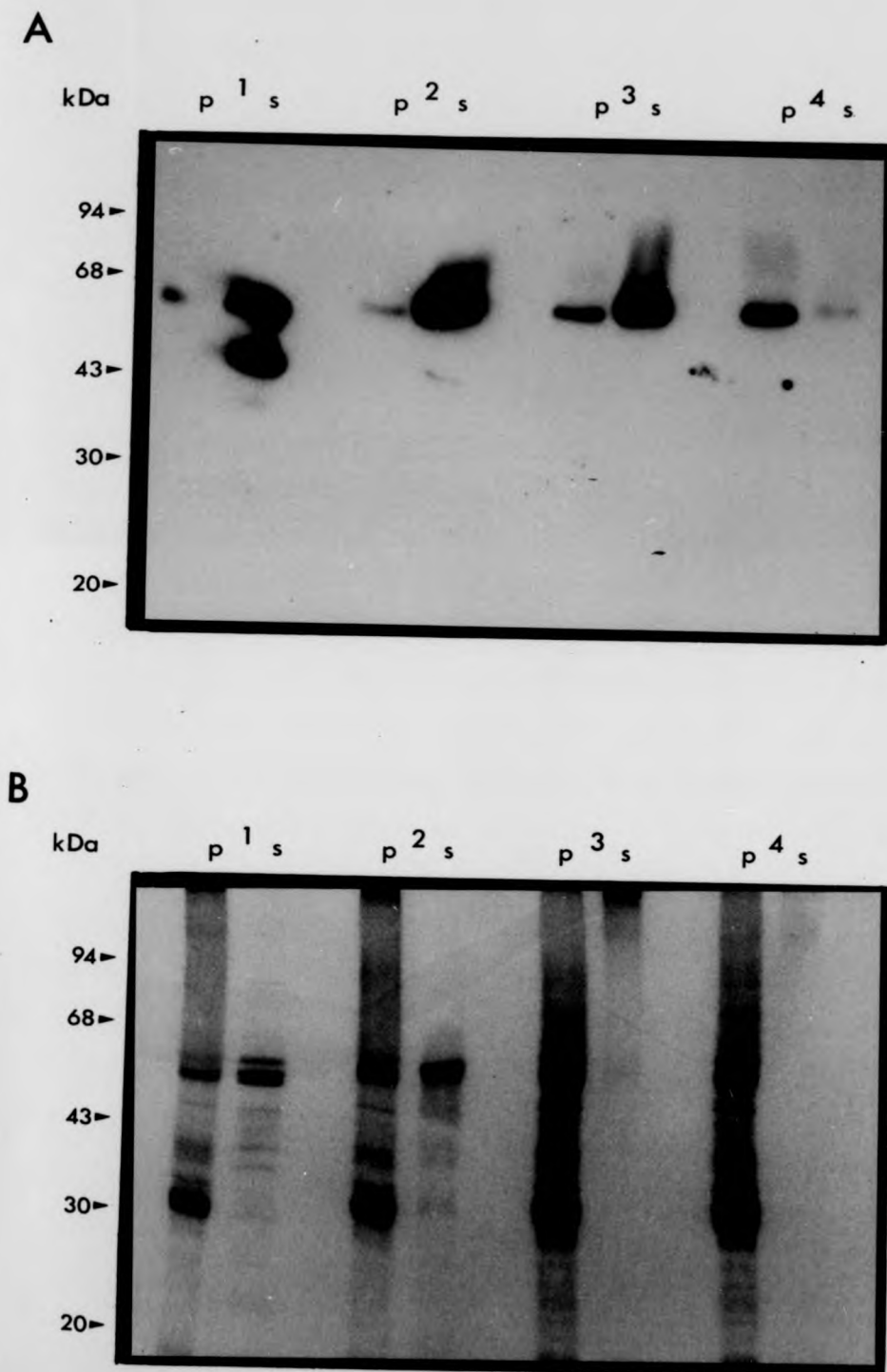




Figure 44. Release of extrinsic proteins from mitochondrial membranes isolated from *Pisum sativum*.

For a description, see legend to Figure 43. The autoradiograph (Panel A) and the Coomassie brilliant blue stained gel (Panel B) are presented, in which lane 1, 5 mM CHAPS; lane 2, 15 mM CHAPS; lane 3, 40 mM CHAPS; lane 5, 70 mM CHAPS, p = membrane fraction and s = solubilized protein fraction. The samples were kindly provided by A.Liddell and C.J. Leaver; molecular weight markers are indicated (see legend to Figure 32).



used by Prasrad and Hallberg (1989) is not sensitive enough to detect the chaperonin in chloroplast fractions.

### 3.3.4 STRUCTURAL AND AMINO ACID SEQUENCE ANALYSIS OF THE CHAPERONIN CLASS

#### 3.3.4.1 Introduction

Chaperonins have been defined as a class of related proteins in plastids, mitochondria and prokaryotes whose proposed role is to mediate protein-protein interactions, required in possibly a whole variety of cellular processes (Hemmingsen *et al.*, 1988). This proposed role raises the question as to what it is that the chaperonins recognize in other proteins, by what means they guide e.g. protein folding and assembly and whether the mode of action is uniform for all chaperonins independent of the cellular process involved. The report of Bradley *et al* (1986) that the wheat Rubisco large and small subunits expressed in *E.coli* fail to assemble into an enzymically active Rubisco oligomer suggests that the bacterial chaperonin is sufficiently different from the chloroplast chaperonin that it is unable to mediate the assembly of a plant Rubisco enzyme. However the observation that enzymically active cyanobacterial Rubisco can be assembled from its expressed subunits in *E.coli* (Gatenby *et al.*, 1985) suggest that the *E.coli* chaperonin is able to mediate the assembly of Rubisco from prokaryotes. Recently this possibility was confirmed by Goloubinoff *et al* (1989), who demonstrated that the assembly of the cyanobacterial Rubisco enzyme is reduced to a low-level in *E.coli* cells that carry a mutation in the chaperonin (i.e. the *groEL*) gene. However they also showed that a mutation in the *groES* gene reduces cyanobacterial Rubisco assembly similarly, while the synthesis of the Rubisco subunits is not affected by the *groE* mutations. Whether a *groES*-like protein is present in chloroplasts is unknown but requires investigation. The possible role of the *groES* protein in protein assembly is discussed later (Section 3.3.5.1). The

observation of Goloubinoff *et al.* (1989) suggests that the *E.coli* chaperonin is capable of replacing at least one function of the cyanobacterial chaperonin, implying that they are functionally highly related. At this stage of the work a number of chaperonin sequences became available, some from unpublished sources. A survey was performed to identify regions that contain large numbers of identical and conserved amino acids in all the available chaperonin sequences. A similar survey was performed to identify conserved regions in the different chaperonin groups i.e. the plastid, the prokaryotic, and the mitochondrial sequences. Possible unique differences between the groups are discussed.

#### 3.3.4.2 The plastid chaperonin group

Four plastid chaperonin amino acid sequences, two of which represent the Rubisco subunit binding protein  $\alpha$  subunit from *Triticum aestivum* (Hemmingsen *et al.*, 1988) and *Brassica napus* one sequence that codes for the  $\beta$  subunit from *Brassica napus* and one sequence that codes for the  $\alpha$  subunit present in the leucoplasts of the endosperm of *Ricinus communis* (Hemmingsen *et al.*, 1988), were compared to determine the percentage amino acid identity. The sequences for the  $\alpha$  and  $\beta$  chaperonins from *B.napus* were kindly provided by S.M. Hemmingsen prior to publication. The three  $\alpha$  chaperonins contain 81-87% identical amino acids (Table 15). Interestingly the  $\alpha$  and the  $\beta$  chaperonins from *B.napus* are only 50% identical; and  $\alpha$  and  $\beta$  are thus as distinct from each other as they are from the prokaryotic chaperonins; the *T.aestivum*  $\alpha$  chaperonin and the *E.coli* chaperonin contain 46% identical amino acids. It is not known whether this difference between the  $\alpha$  and  $\beta$  chaperonins has any functional significance. The plastid chaperonin sequences show a high degree of identical and conserved amino acids distributed along the entire length of the polypeptide chains (Fig. 45). Certain regions show a very high percentage identity e.g. between residue 22 and 68 (72%), between 73 and 102 (75%) and between 366 and 422 (78%). The latter region is of interest since it is this region in



Table 15. Percentage identical amino acids of the chaperonin proteins

The numbers show percentage identical amino acids. The algorithm used was described by Needleman & Wunsch (1970), and the Ubitary Matrix which scores only identities and mismatches (von Heijne, 1987).

Table 15. Percentage identical amino acids of the chaperonin sequences

Symbol	Chaperonin											
M1	mitochondrion ( <i>H.sapiens</i> )	100										
M2	mitochondrion ( <i>S.cerevisiae</i> )	71	100									
P1	<i>M.leprae</i>	47	47	100								
P2	<i>M.tuberculosis</i>	45	47	95	100							
P3	<i>E.coli</i>	50	53	58	59	100						
P4	<i>C.brunetii</i>	52	51	60	61	76	100					
P5	<i>A.nidulans</i>	48	50	64	64	78	59	100				
C1	alpha ( <i>R.communis</i> )	45	45	51	51	49	48	56	100			
C2	alpha ( <i>T.aestivum</i> )	43	41	49	49	46	48	54	84	100		
C3	alpha ( <i>B.napus</i> )	44	42	53	51	48	46	58	87	81	100	
C4	Beta ( <i>B.napus</i> )	43	46	53	51	52	52	55	52	51	50	100
		M1	M2	P1	P2	P3	P4	P5	C1	C2	C3	C4

The derived amino acid sequence of the plastid chaperonin  $\alpha$  subunit from *Triticum aestivum* (C2) and from *Brassica napus* (C3), the  $\beta$  subunit from *Brassica napus* (C4) and the  $\alpha$  subunit from *Ricinus communis* endosperm (C1) were compared and a consensus sequence of identical amino acids was determined (CON). The sequences for the  $\alpha$  and the  $\beta$  subunits of *B.napus* were kindly provided by S.M. Hemmingsen prior to publication. The one letter amino acid code is presented (IUPAC-IUB, 1970; 1984) and regions containing > 71% identical amino acid residues are boxed. Vertical lines show identical amino acids, and dots indicate gaps. The Needleman and Wunsch (1970) algorithm was used.

154/3

154/4

C1 APLLIIAEDVTGEALASLVVNKMRGILNVAAIKVPGFGERRKALLQDIAI  
 C2 CPLFIVAEDITGEALATLVVNKL RGIINVAAIKAPSFGERRKAVLQDIAI  
 C3 APLLIIAEDVTGEALATLVVNKL RGV LNVVAVKAPGFGERRKAMLQDIAI  
 C4 YPILIIAEDIEQEALATLVVNKL RGT LKIAALKAPGFGERKSQYLDIAI  
 CON -P-LIIAED---EALA-LVVNKL R G---V-A-K-P-FGER----L-DIAI  
 -----+-----+-----+-----+-----+ 300

LTGAEFQASDLG L LVENTSVEQLGIARKVTITKDSTTLIADAASKDELQA  
 VTGAEYLAKDLG L LVENATVDQLGTARKITIHQTTTLIADAASKDEIQA  
 LTGPSTT.LDMG L LVENTTIDQLGIARKVTISKDSTTLIADAASKETLQA  
 LTGATVIREEVGLSLDKAGKEVLGHAAKVVLTKETSTIVGDGSTQDAVQK  
 -TG-----D-GL--E-----LG-A-KV-----T-I-D----D--Q-  
 -----+-----+-----+-----+-----+ 350

RIAQLKRELAETDSVYDSEKLAERIAKL SGGVAVIKVGAATETELED RKL  
 RVAQLKKELSETDSIYDSEKLAERIAKL SGGVAVIKVGATTETELED RQL  
 RISQLKKELSETDSVYDSEKLAERMPKLPGGVAVIKVGAATETELED RKL  
 RVTQIKNLIEQAEQDYKEKLNERIAKL SGGVAVIQVGAQTETELKEKKL  
 R--Q-K-----D--YD-EKL-ER--KL-GGVAVI-VGA-TETEL-D--L  
 -----+-----+-----+-----+-----+ 400

RIEDAKNATFAAIEEGIVPGGGAALVHLSTVVPAINGE.DDADERLGAD.  
 RIEDAKNATFAAIEEGIVPGGGAAYVHLSTYVPAIKETIEDHDERLGADI  
 RIEDAKNATFAAIEEGIVPGGGATLVHLSTVIPAIKEKLEDADERLGADI  
 RVEDALNATKAAVEEGIVVGGGCTLVCLASKVD A IKAALDNDEEKVGADI  
 RIEDA-NAT-AAIEEGIV-GGG--V-L---V-AI-----DE--GAD-  
 -----+-----+-----+-----+-----+ 450

...RHWYPTSLIAQNAGIEGEVVVEKVKAREW.EIGYNAMTDKYENLVEA  
 IQKALQAPASLIANNAGVEGEVVIEKIKESEW.EMGYNAMTDKYENLIES  
 VQKAL.VARSLIAQNAGIGGEVVVEKIMFSEW.EIGYNAMTDTYENLLEA  
 VKRALSYPKLIAKNAGVNGSVVSEKVLSDNVKFGYNAATGKYEVSMAA  
 -----LIA-NAG--G-VV-EK---E----GYNA-T--YE-----  
 -----+-----+-----+-----+-----+ 500

C1 GVIDPAKVTRCALQNS  
 |||||  
 C2 GVIDPAKVTRCALQNAASVSGMVLTTQAIIVVERPKPKPKVAEPAEGQLSV  
 |||||  
 C3 GVIDPAKVTRCALQNAASVAGMVLTTQAIIVVDKPKPKAPTAAFPQGLMV  
 |||||  
 C4 GIIDPTKVVRCCLEHAASVAKTFLMSDCVVVEIKEPEPVP  
 CON GVIDP-KV-RC-L-----  
 -----+-----+-----+-----+ 550

which the reversed nucleotide binding site sequence in the wheat  $\alpha$  chaperonin polypeptide has been identified (see Section 3.2.5.4). Secondary structure predictions of the chloroplast chaperonins show that although the  $\alpha$  and  $\beta$  chaperonins are only 50% identical, at the a.a. level their structures are remarkably similar (Fig. 46).

#### 3.3.4.3 The prokaryotic chaperonin group

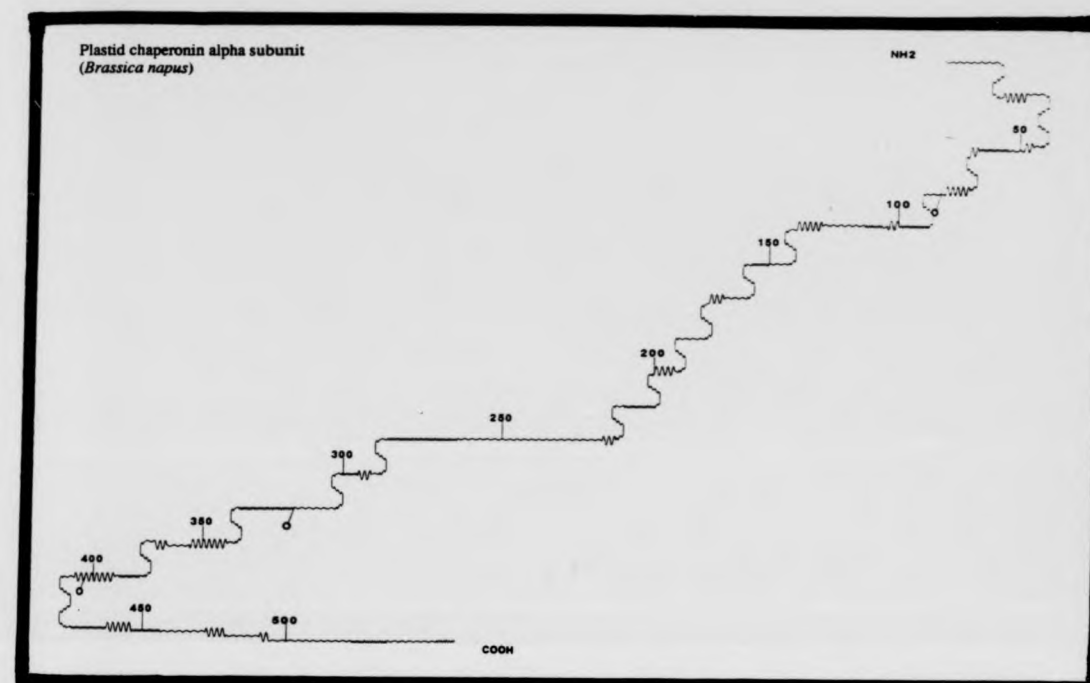
Comparison of five prokaryotic chaperonin sequences from *Mycobacterium leprae* (Mehra *et al.*, 1986), *Mycobacterium tuberculosis* (Shinnick, 1987), *Escherichia coli* (Hemmingsen *et al.*, 1988) *Coxiella burnetii* (Vodkin & Williams, 1988) and *Anacystis nidulans* (Cozens & Walker, 1987) shows that 95% of the mycobacterial chaperonin amino acids are identical whereas the *E.coli* and *C.burnetii* sequences shown 76% identity (Table 15). As a group, the prokaryotic chaperonins shown a slightly higher percentage of identity (between 60% and 70%) than the plastid  $\alpha$  and  $\beta$  chaperonins (50%), indicating that the  $\alpha$  and  $\beta$  subunits of the plastid chaperonins are less related than the chaperonins from different prokaryotes. Long highly conserved regions of amino acids are found when the prokaryotic sequences are aligned (Fig. 47). The region from residue 18 to 124 contains 86% identical amino acids, while between residue 358 and 418, 82% of amino acids are identical. The latter region includes only four chaperonin sequences since only the first 300 amino acids of the *A.nidulans* chaperonin are known. The identified regions are also highly conserved in the plastid chaperonins.

The cyanobacterial chaperonin seems more closely related to the other prokaryotic chaperonins (78% identity to *E.coli*) than to the plastid chaperonins (about 50% identity). However, when conserved amino acid substitutions are also considered, the similarity with the prokaryotic and plastid chaperonins is very much the same (approximately 80%). An indication that the cyanobacterial chaperonin is more similar to the plastid chaperonin becomes evident when the predicted secondary structures are compared. A striking resemblance is observed between the

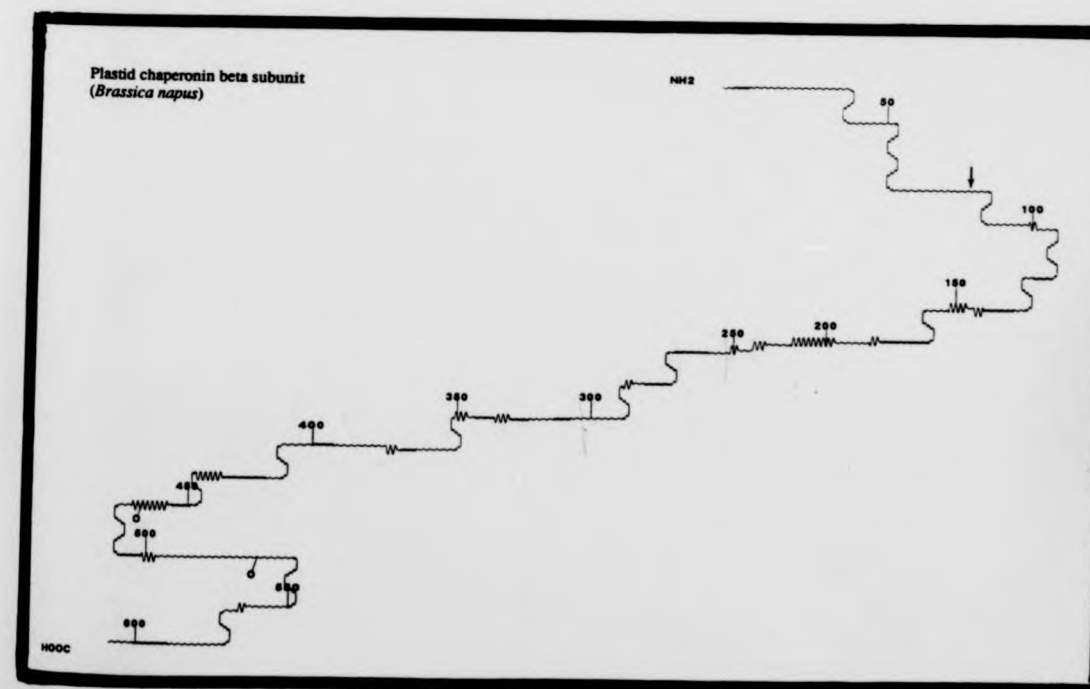
Figure 46. Secondary structure prediction of the *Brassica napus* chloroplast chaperonins

The secondary structure predictions (Chou & Fasman, 1978) for the  $\alpha$  (Panel A) and  $\beta$  subunits (Panel B) are presented as two dimensional plots. The programme 'plotstructure' from the GCG software package (Devereux *et al.*, 1984) was used to produce the plots. Helices are shown with a sine wave,  $\beta$  sheets with a sharp tooth wave, turns with 90 degree turns, and random coils with a dull saw-tooth wave. Putative glycosylation sites are indicated (--O). The arrow (Panel B) indicates the probable cleavage site to remove the presequence and produce the mature  $\beta$  subunit.

A



B



Derived amino acid sequences of the *Mycobacterium leprae* 65 kDa common antigen (P1; Mehra et al., 1986), the *Mycobacterium tuberculosis* 65 kDa common antigen (P2; Shinnick, 1987), the *Escherichia coli* groEL protein (P3; Hemmingsen et al., 1988), the *Coxiella burnetii* major antigen (P4; Vodkin & Williams, 1988) and the *Anacystis nidulans* URF3 (P5; Cozens & Walker, 1987) are compared and a consensus sequences (CON) showing amino acids that are identical in at least four of the sequences (identity 80% or more). Highly conserved regions are boxed. Vertical lines show identical amino acids, and dots indicate gaps. The algorithm described by Needleman & Wunsch (1970) was used.

155/3



P1 SGYFVTD AERQEAVLEEPYILLVSSKVSTVKDLLP LLEKVIQAGKSLII  
P2 SGYFVTD AERQEAVLEDPYILLVSSKVSTVKDLLP LLEKVIQAGKPLII  
P3 SPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAVAKAGKPLII  
P4 SPYFINNQNM SAELENPFILLVDKKISNIRELIPLLENVAKSGRPLVI  
P5 SPYFATDTERMEAVFDEPFILITDKKIGLVQDLVPVLEQVARAGRPLVII  
CON SPYFI---E---A-LE-PFILL-DKKIS----L-P-LE-VA-AG-PLII  
-----+-----+-----+-----+-----+ 250

AEDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAMLQDMAILTGAQV  
AEDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAMLQDMAILTGGQV  
AEDVEGEALATAVVNTIRGIVKVA V KAPGFGDRRKAMLQDIATLTGGTV  
AEDIEGEALATLVVNNIRGVV KVA V KAPGFGDRRKAMLQDIAVLTGGKV  
AEDIEKEALATLVVNRLRGVLNVA V KAPGFGDRRKAMLEDIAVLTGGQL  
AED-EGEALATLVVN-IRGV-KVA V KAPGFGDRRKAMLQDIAVLTGG-V  
-----+-----+-----+-----+-----+ 300

ISEEVGLTLENTDLSLLGKARKVVM TKDETTIVEGAGDTDAIAGRVAQIR  
ISEEVGLTLENADLSLLGKARKVVVTDKETTIVEGAGDTDAIAGRVAQIR  
ISEEIGMELEKATLEDLGQAKRVVINKD TTTIIDGVGEEAAIQGRVAQIR  
ISEEVGLSLEAASLDDLGS AKRVVVT KD DTTIIDGSGDAGDIKNRVEQIR  
I  
ISEEVGL-LE---L--LG-A--VV--KD-TT IIDG-GD---I--RV-QIR  
-----+-----+-----+-----+-----+ 350

TEIENS DSDYDREKLQERLAKLAGGVAVIKAGAATEVELKERKHRIEDAV  
QEIENS DSDYDREKLQERLAKLAGGVAVIKAGAATEVELKERKHRIEDAV  
QQIEEATSDYDREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDAL  
KEIENS SSDYDKEKLQERLAKLAGGVAVIKVGAATEVEMKEKKARVEDAL  
--IE--SDYD-EKLQER-AKLAGGVAVIK-GAATEVEMKE-K-RVEDA-  
-----+-----+-----+-----+-----+ 400

P1 RNAKAAVEEGIVAGGGVTLLQAAPALDKLKTGDEA.TGANIVKVALEAP  
P2 RNAKAAVEEGIVAGGGVTLLQAAPTLDLKLKLEGDEA.TGANIVKVALEAP  
P3 HATRAAVEEGVVAGGGVALIRVASKLADLRGQNEQNVGIKVALRAMEAP  
P4 HATRAAVEEGVPPGGGVALIRVLKSLDSVEVENEDQRVGVEIARRAMAYP  
CON -----AAVEEGVV-GGGV-L-----L-----ED--G--I--AM--P  
-----+-----+-----+-----+-----+ 450

LKQIAFNSEMPGVVAAEKVRNLSVGH.GLNAATGEYEDLLKAGVADPVKV  
LQKIAFNSEMPGVVAAEKVRNLPAGH.GLNAQTGVYEDLLAAGVADPVKV  
LRQIVLNCGEPSVVANTVKG.GDGNYGYNAAATEEYGNMIDMGILDPTKV  
LSQIVKNTGVQAAVVADKVLNHNKDVNYGYNAATGEYGDMIEMGILDPTKV  
L-QI--N-G---VVA--V-----G-NAAT-EY--M---GI-DP-KV  
-----+-----+-----+-----+-----+ 500

TRSAQNAASIAGLFLTTEAVVADKPEKTAA..PASDPTGGMGGMDF  
TRSAQNAASIAGLFLTTEAVVADKPEKEKASVPG.G..GDMGGMDF  
TRSAQYAASVAGLMITTECMVTDLPKNDAA DLGAAGMGGMGGMGMM  
TRTAQNAASIAGLMITTECMVTEAPKKKEESMPGGGDMGGMGGMGGMG  
TR-ALQ-AASIAGL--TTE--V-D-P-----GGMGGM-----  
-----+-----+-----+-----+-----+ 550

P4 MM  
-- 552

Table 16. Percentage conserved amino acids of the chaperonin proteins

The numbers show percentages of similarity and were determined using the algorithms described by Needleman & Wunsch (1970) and Sellers (1974); positive scores were given to evolutionarily frequent, conservative replacements of one amino acid for a chemically similar one.

Table 16. Percentage conserved amino acids of the chaperonin sequences

Symbol	Chaperonin										
M1	mitochondrion ( <i>H.sapiens</i> )	100									
M2	mitochondrion ( <i>S.cerevisiae</i> )	72	100								
P1	<i>M.leprae</i>	66	68	100							
P2	<i>M.tuberculosis</i>	66	68	97	100						
P3	<i>E.coli</i>	69	70	75	75	100					
P4	<i>C.brunetti</i>	71	70	76	76	85	100				
P5	<i>A.nidulans</i>	68	69	81	81	78	78	100			
C1	alpha ( <i>R.communis</i> )	65	65	70	70	70	72	76	100		
C2	alpha ( <i>T.aestivum</i> )	65	64	70	71	71	70	76	92	100	
C3	alpha ( <i>B.napus</i> )	64	65	71	70	70	70	72/	93	91	100
C4	beta ( <i>B.napus</i> )	64	64	69	68	73	73	77	71	70	70 100
		M1	M2	P1	P2	P3	P4	P5	C1	C2	C3 C4

cyanobacterial and the chloroplastic chaperonin structures, which are different from the structure of the *E.coli* chaperonin (Fig. 48).

Comparison of the prokaryotic chaperonin secondary structures reveals that a unique conformation is present in the *C.burnetii* chaperonin i.e. random coil/turn/random coil (Fig. 49C). This unique conformation is found in every region where turns are predicted. In all the other prokaryotic chaperonin structures turns are flanked by  $\beta$ -sheet conformations rather than random coils. The significance of the unique conformation in the *C.burnetii* chaperonin secondary structure is not known. All prokaryotic chaperonins contain a screw-like structure at the carboxyterminus; which is not observed in the plastid chaperonins. In this screw-like structure, only the *C.burnetii* chaperonin has the unique random coil/turn/random coil conformation. The screw-like structure is the result of the (GGM)<sub>n</sub>-motif, where n = 1 to 4. The (GGM)-motif is shorter in the *M.tuberculosis* sequence (8 residues) than in the *C.burnetii* sequence (16 residues) and this is apparent in the secondary structure (cf. Fig. 47; Fig. 48; Fig. 49). Whether this screw-like structure, typical for the prokaryotic chaperonins, is of functional significance remains to be seen.

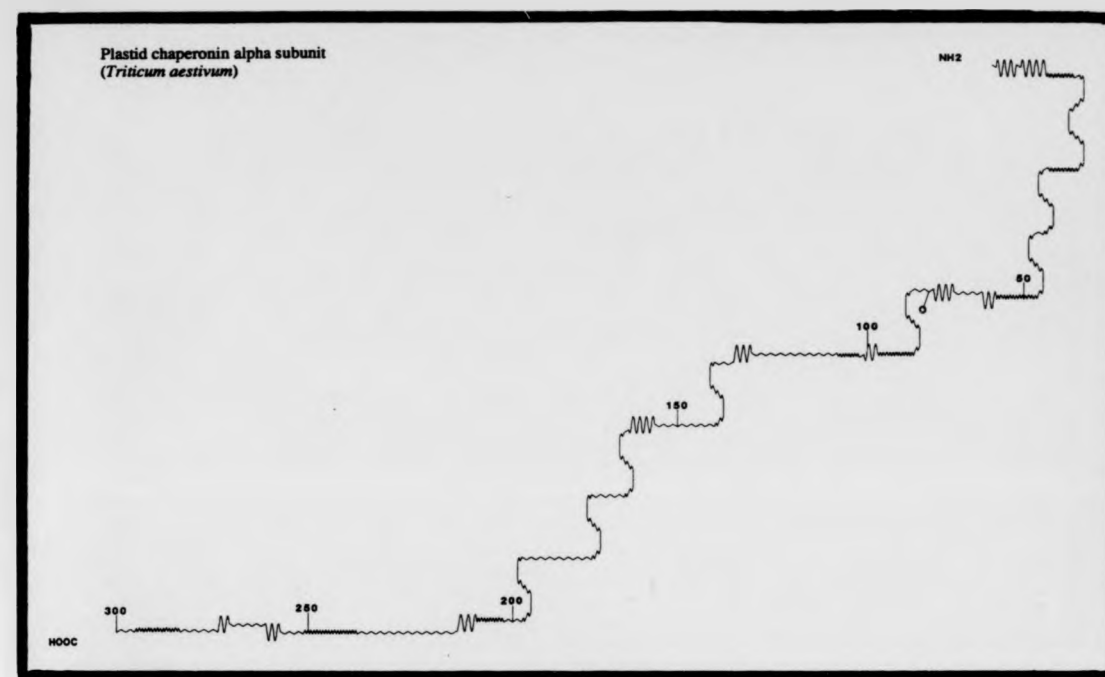
#### 3.3.4.4 The mitochondrial chaperonin group

The two mitochondrial chaperonin sequences from *Homo sapiens* (Jindal *et al.*, 1989) and *Saccharomyces cerevisiae* (Reading *et al.*, 1989) are 71% identical at the amino acid level (Table 15) and are thus more highly conserved than most of the prokaryotic chaperonin sequences (between 50% and 60% identity). A number of short regions of 9 to 13 amino acids are identical in both the mitochondrial sequences (Fig. 50). The mitochondrial chaperonins, like the prokaryotic ones, contain the (GGM)<sub>n</sub>-motif at the carboxyterminus. The secondary structures of the yeast and the human mitochondrial chaperonins are different and seem to be mirror images (Fig. 51). It was noticed that the human chaperonin secondary structure resembles that of the higher

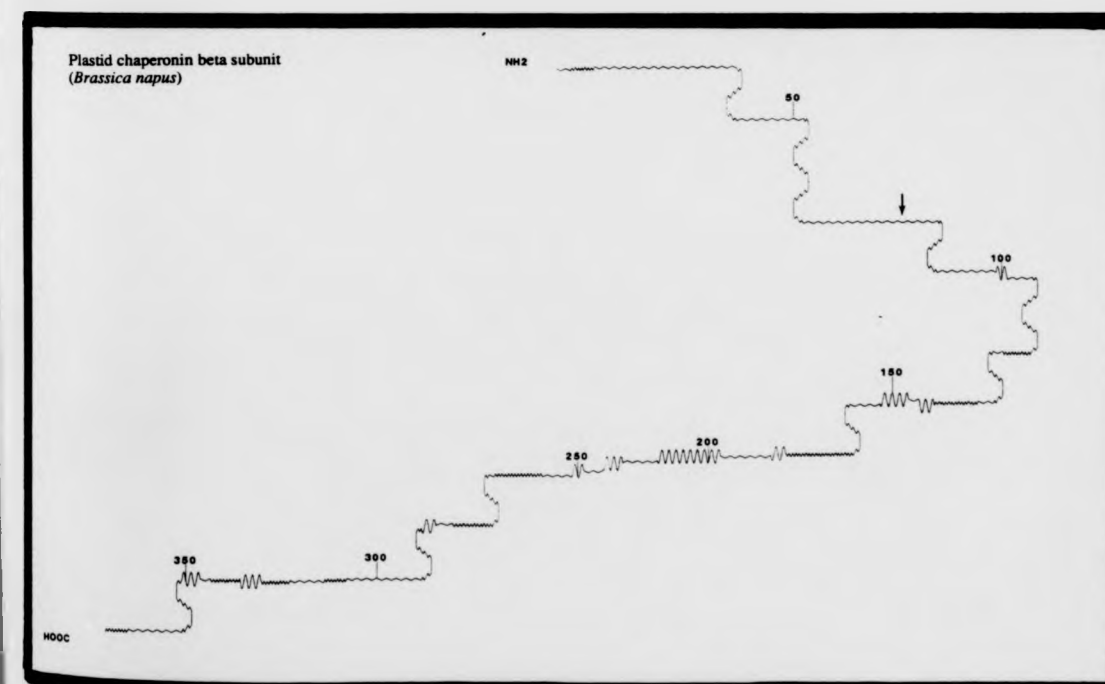
Figure 48. Comparison of secondary structure predictions of chloroplast and prokaryotic chaperonins

The secondary structure predictions according to Chou & Fasman (1978) are presented in the form of a two dimensional plot. For details see the legend to Figure 46. A putative glycosylation site (—O) is indicated in Panel A and the arrow in Panel B indicates the possible cleavage site of the *B.napus*  $\beta$  subunit precursor.

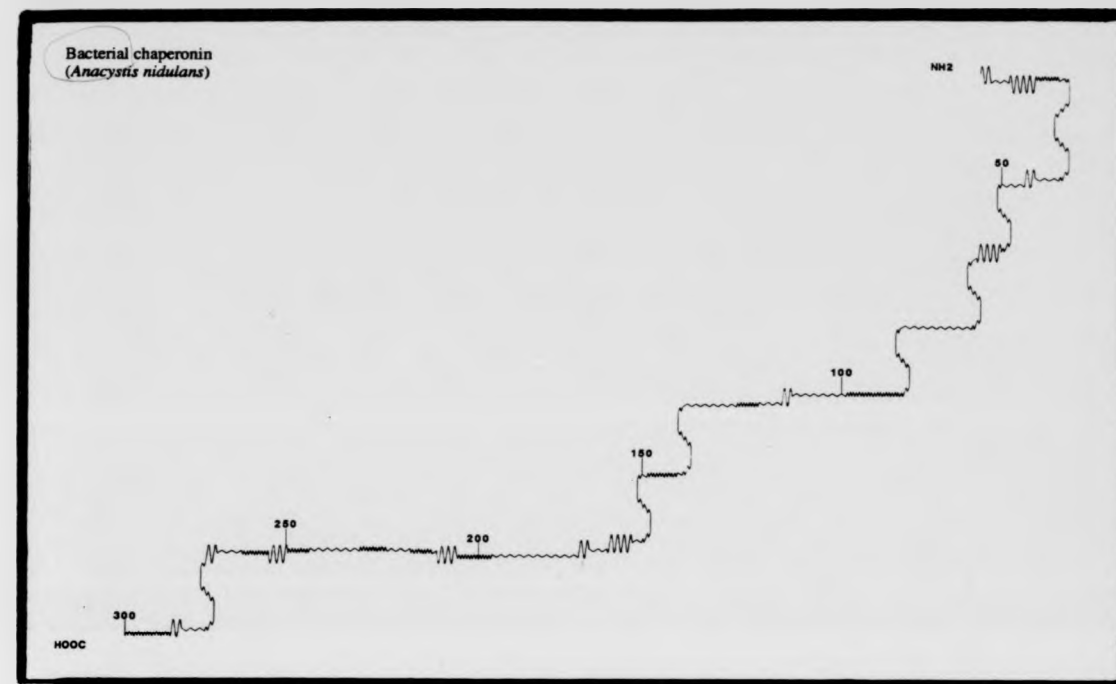
A



B



C



D

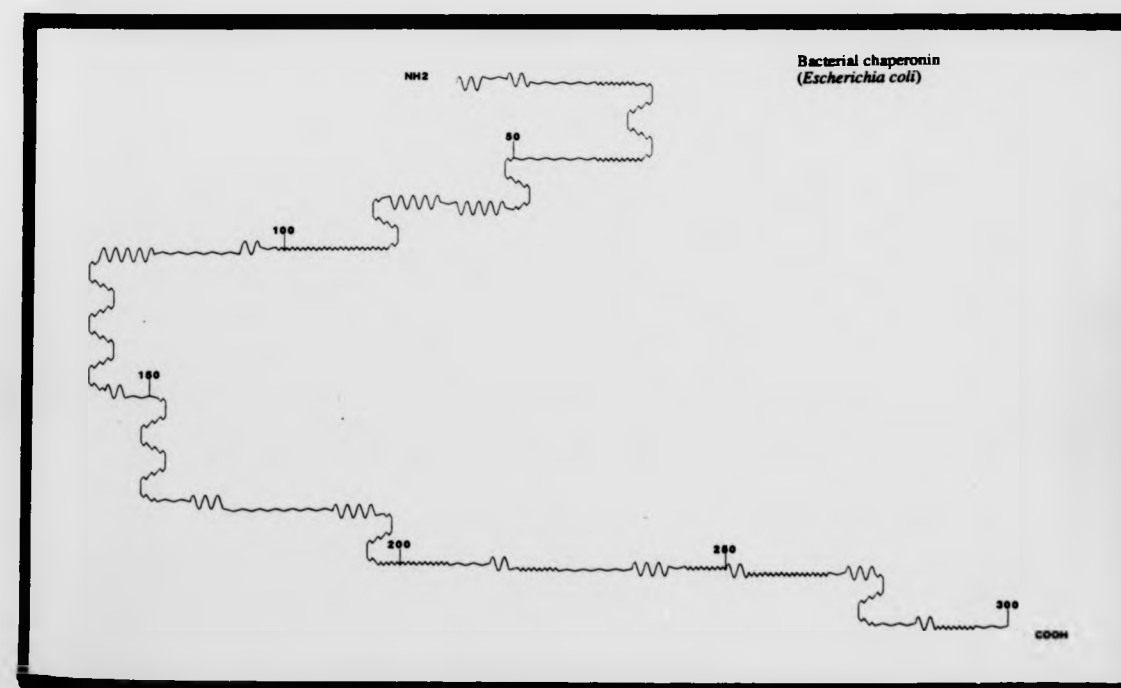
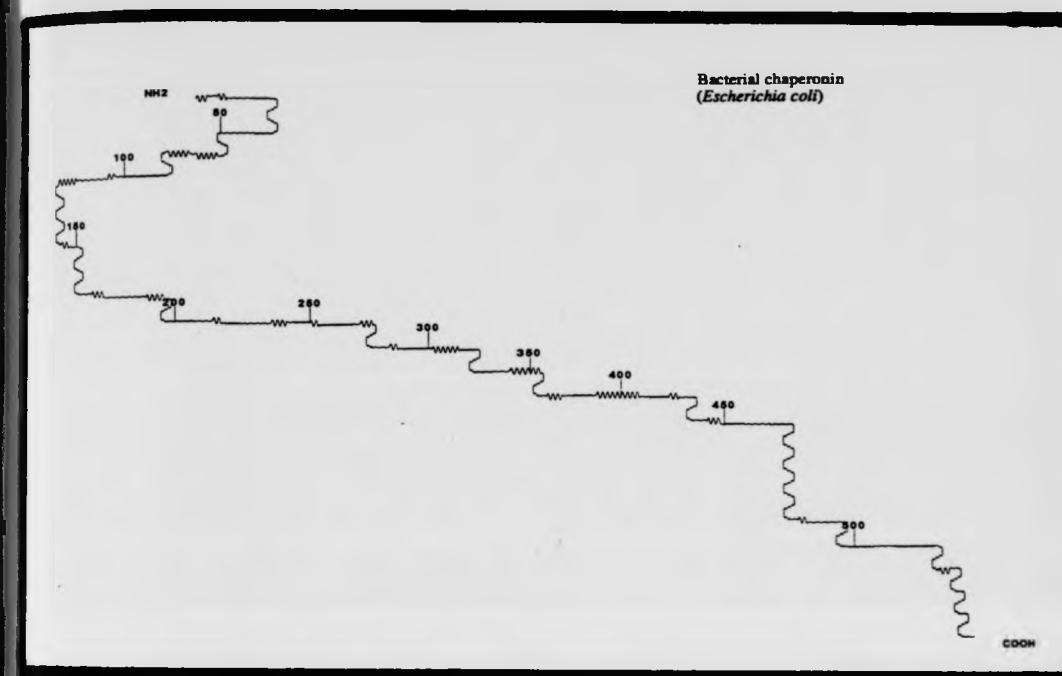


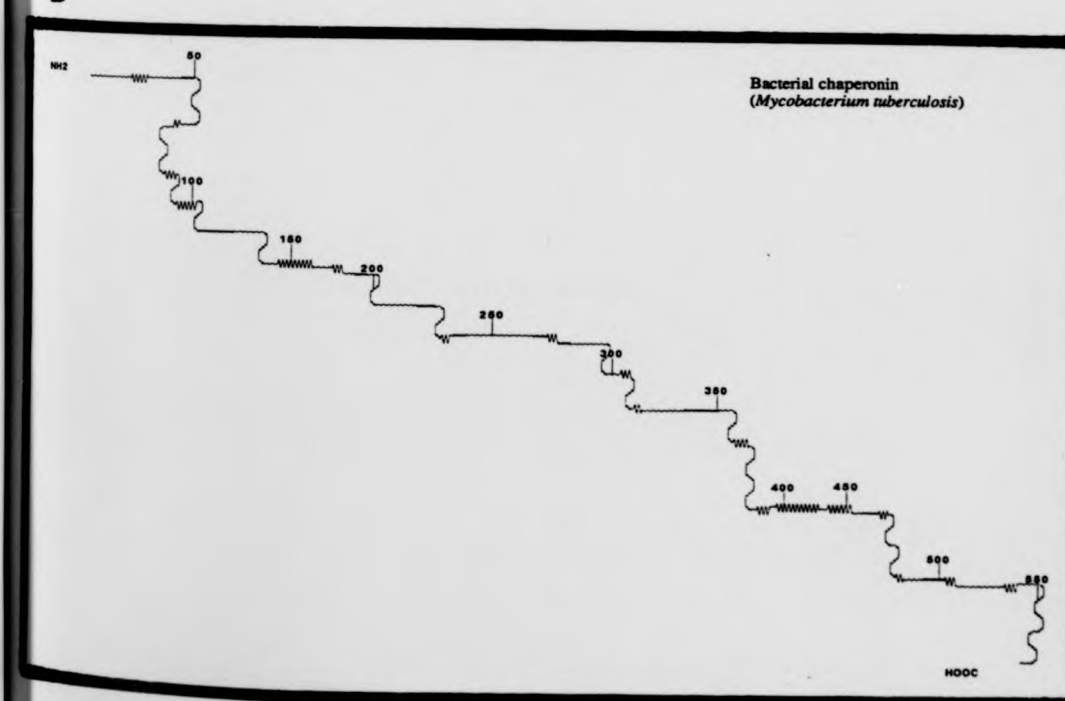
Figure 49. Secondary structure predictions of the prokaryotic chaperonins

The secondary structure predictions (Chou & Fasman, 1978) are shown as two dimensional plots. For details see the legend to Figure 46. Putative glycosylation sites are indicated (—O).

A



B





C

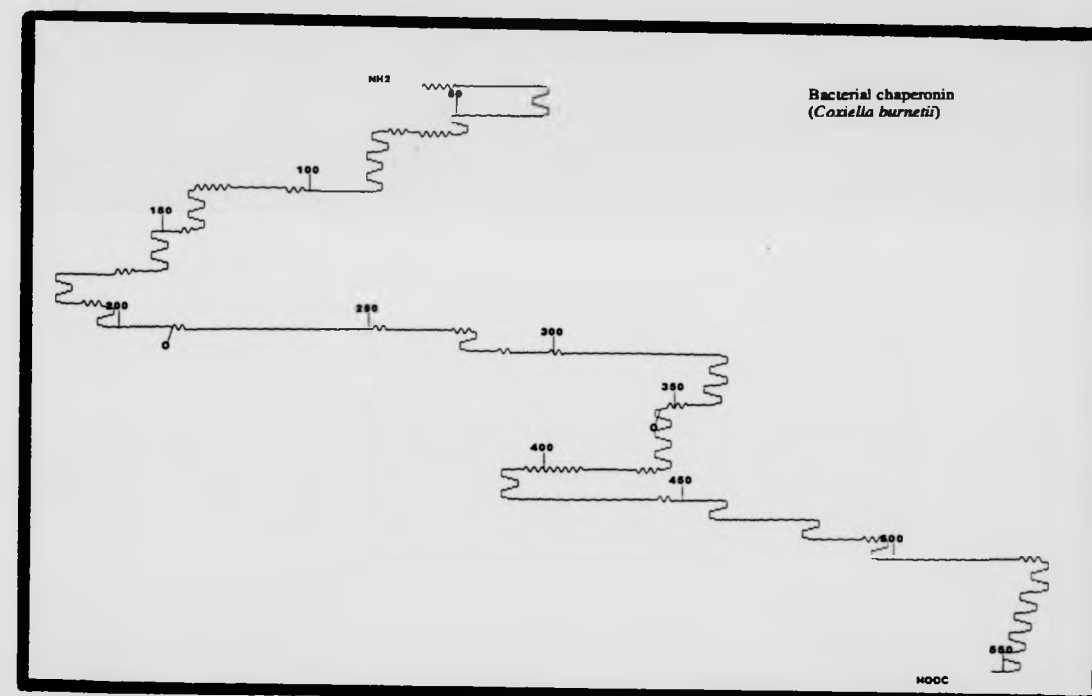


Figure 50. Amino acid sequence comparison of the mitochondrial chaperonins

The derived amino acid sequences of the *Homo sapiens* mitochondrial chaperonin (M1; Jindal et al., 1989) and the *Saccharomyces cerevisiae* mitochondrial chaperonin (M2; Reading et al., 1989) are compared and a consensus (CON) sequence of identical amino acids is shown. Highly conserved regions with identical amino acids residues are boxed. Vertical lines show identities and dots indicate gaps. The algorithm used was described by Needleman & Wunsch (1970)

```

M1  MLRLPTVFRQMRPVSRVLAPHLTRAYA..KDVKFGADARALMLQGVDLLA
M2      MLRSSVVRSRATLRPLLRAYSSHKDLKFGVEGRASLLKGVETLA
CON  -----L-P-L-RAY---KD-KFG---RA--L-GV--LA  50
      +-----+-----+-----+-----+

DAVAVTMGPKGRTVII EQSWGSPKVT KDGVTVAKSIDLKDKYKNIGAKLV
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
EAVAATLPGKGRNVLI EQAFGPPKITDKGVTKAVSIVLKDKFENMGAKLL
      +-----+-----+-----+-----+-----+ 100

QDVANNTNEEAGDGT TTTATV LARSIAKEGF EKISKGANPVEIRRGV MLAV
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QEVASKTNYAAGDGT SATVLGRAI FTESVKNVAAGYNPMDLR RGSQVAV
      +-----+-----+-----+-----+-----+ 150

DAVIAELKKQSKPVTTPEEIAQVATISANGDKEIGNIISDAMKKVGRKGV
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ERVIEFLSANKKEITTSEEIAQVATISANGDSHVGKLLASAMEKVGKEGV
      +-----+-----+-----+-----+-----+ 200
      --VI--L---K--TT-EEIAQVATISANGD--G-----AM-KVG--GV

ITVKDGKTLNDELEII EGMKFDRGYISPYFINTSKGQKCEFQDAYVLLSE
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ITIREGRTLEDELEVTEGMRFD RGFISKY NITDPKSDKVEFEKPLLLSE
      +-----+-----+-----+-----+-----+ 250
      IT---G-TL-DELE--EGM-FDRG-IS-Y-I---K--K-EF---LLSE

KKISSIQSIVPALEIANAH RKPLV IIAEDVDGEALSTLVNRLKVGLQVV
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
KKISSIQDILPALEISNQSR RPLI IIAEDVDGEALAACILNNLRGQVKVC
      +-----+-----+-----+-----+-----+ 300
      KRISSIQ-I-PALEI-N--R-PL-IIAEDVDGEAL---LN-L-----V-

```

AVKAP-FGDNRKN---D-A--TGG-VF-EE-L-L--E-----LG-----  
 -----+-----+-----+-----+-----+-----+-----+ 350

[illegible]

DGVAVLKVGGTSDVEVNEKKDRVTDALNATRAAVEEGIVLGGGCALLRCI  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
GGVAVIRVGGASEVEVGEKKDRYDDALFATRAAVEEGI LPGGGTALVKAS  
  
-GVAV--VGG-S-VEV-EKKDR--DAL-**ATRAAVEEGI**-GGG-AL----

450

PALDSLTPANEDQKIGIEIIKRTLKIPAMTIAKNAGVEGSLIVEKIMQSS  
 || | || | || | || | || | || | || | || |  
 RVLDEVVVDNFDQKLGVDIIRKAITRPAKQIIENAGEEGSVIIGKLIDEY  
 --LD-----N-DQK-G--II-----PA--I--NAG-EGS-I--K-----  
 -----+-----+-----+-----+-----+-----+ 500

SE...VGYDAMAGDFVNMEKGIIDPTKVVRTALLDAAGVASLLTTAEVV  
||| | |||| | | |||| |  
GDDFAKGYDASKSEYTDMLATGIIDPFKVVRSGLVDSGVASLLATTEVA

-----GYDA-----M---GIIDP-KVVR--L-DA-GVASLL-T-EV-

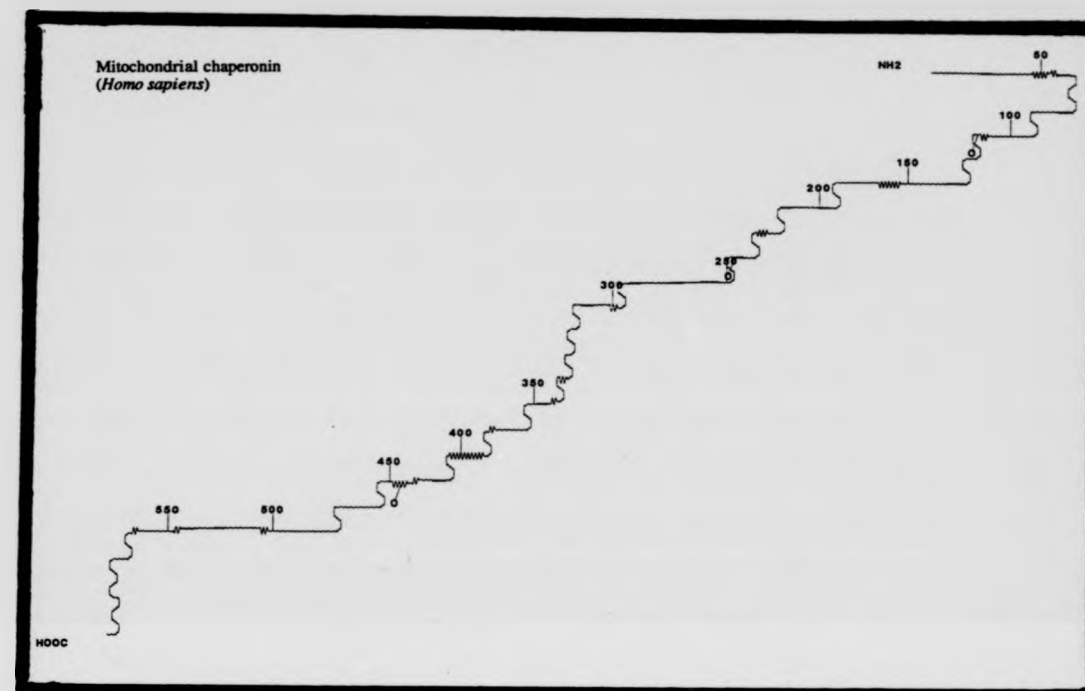
+-----+-----+-----+-----+-----+-----+ 550

VTEIPKEEKDPGMGAMGGMGGGMGGGMF  
| | | | | | | | | |  
IVDAPEPPAAAGAGGMPGGMPGMPGMM.  
-----P-----G-G-M-G---GM-G-M-  
-----+-----+----- 578

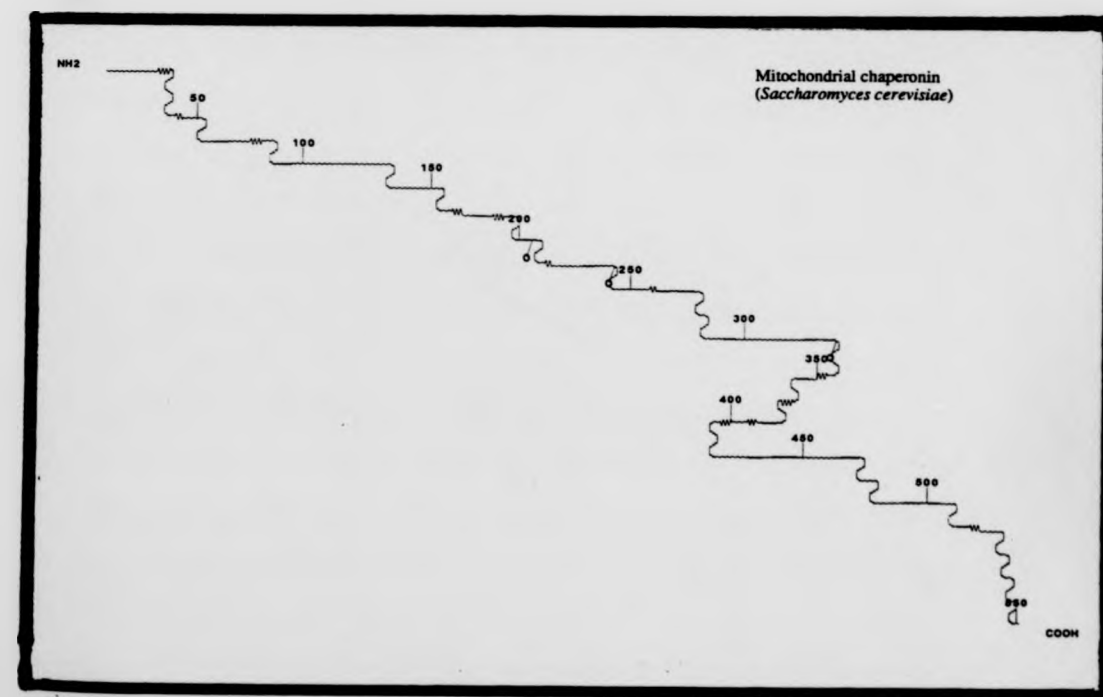
Figure 51. Secondary structure predictions of the mitochondrial chaperonins

The secondary structure predictions (Chou & Fasman, 1970) are presented as two dimensional plots. For details see the legend to Figure 46. Possible glycosylation sites are indicated (—O).

A



B



eukaryotic plant chaperonins while the yeast chaperonin secondary structure is more similar to the prokaryotic chaperonin structures.

It is concluded that there are no significant differences of highly conserved regions of amino acids between the different chaperonin groups, apart from the carboxyterminal (GGM)<sub>n</sub>-motif which is typical of prokaryotic and mitochondrial chaperonins. However in terms of secondary structure, the prokaryotic and plastid chaperonins are different, which might reflect their functional differences. The cyanobacterial chaperonin structure resembles that of the plastid chaperonins, and, since the *E.coli* chaperonin is able to mediate the assembly of cyanobacterial Rubisco but not plant Rubisco, this suggests that possibly other factors might be involved in the action of the chaperonins such as a *groES*-like protein.

#### 3.3.4.5 The chaperonin class

Since conserved regions of amino acids are apparent between the different chaperonin groups all eleven chaperonin sequences were aligned to identify conserved regions present in all known chaperonins. A consensus sequence was determined for eleven chaperonins and represents 80% similarity (Fig. 52). A total of six highly conserved regions were identified; four of these are short sequences of 12 to 20 amino acids, with an identity of > 90% occurring at positions 110-122; 276-287; 327-346 and 356-368. The two other highly conserved regions are between positions 100 and 200, and between positions 440 and 505 (Fig. 52). The latter region is of great interest since the reversed dinucleotide binding site, identified in the wheat plastid chaperonin  $\alpha$  subunit sequence, occurs in this area. The fact that this region is so highly conserved in the chaperonin sequences suggests it is likely to be of functional significance. The secondary structure, in this region of the *E.coli* chaperonin sequences, is predicted as an  $\beta$ - $\alpha$ - $\beta$  fold and is strikingly identical to the reported  $\beta$ - $\alpha$ - $\beta$  conformation for a known dinucleotide binding site (Wieringa *et al.*, 1985). The structure of this region in the plastid chaperonins is a  $\beta$ - $\alpha$ -turn- $\beta$  conformation, a slight variant (Fig. 53). It

Figure 52. Amino acid sequence comparison of the eleven chaperonins

A consensus sequence of >80% identical amino acid residues was determined (CON). Symbols: M1, *H.sapiens*; M2, *S.cerevisiae*; P1, *M.leprae*; P2, *M.tuberculosis*; P3, *E.coli*; P4, *C.burnetii*; P5 *A.nidulans*; C1, *R.communis* ( $\alpha$ ); C2, *T.aestivum* ( $\alpha$ ); C3, *B.napus* ( $\alpha$ ) and C4, *B.napus* ( $\beta$ ). For detailed information see the legends to Figure 45, Figure 47 and Figure 50. Highly conserved regions are boxed. Dots indicate gaps. The algorithm used was described by Needleman & Wunsch (1970)

```

C4  PAPQFLIRGASPLRFTSLEKMASTFTATSSIGSMVAPNAHKTDKKLMNK
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
50

M1      MLRLPTVFRQMRPVSRLAPHLTRAYAKDVKFGADARAL..MLQG
M2      MLRSSVVRSRATLRPLLRAYSSHKDLKFGVEGRA..SLLKG
P1      MAKTIAYDEEARR..GLERG
P2      MAKTIAYDEEARR..GLERG
P3      MAAKDVKFGNDARVKML..RG
P4      MAAKVLKFSHEVLHAMS..RG
P5      MAKRIIYNENAR..RALEKG
C1      RTALQSG
C2      GADAKEIAFDQKSRA..ALQAG
C3      RRFSVRANVKEISFDQSSRA..ALQAG
C4      LSSSSFGRRQNVFPKLRRSS..PA...VVCAAKELHFNKDGTITRKLQAG
CON  -----AK---F---R---L---G
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
100

VDLLADAVAVTMGPKGRVTVIIEQSWGSPKVTKDGVTVAKSIDLKDKYKNI
VETLAEAVAATLPKGKGRNVLEQAFGPPKITDKGVTKAVSIVLKDKFENM
LNSLADAVKVTLPKGKGRNVVLEKKWGAPTITNDGVSIAKEIELEDPYEKI
LNALADAVKVTLPKGKGRNVVLEKKWGAPTITNDGVSIAKEIELEDPYEKI
VNVLADAVKVTLPKGKGRNVVLDKSFAGAPTITKDGVSVAKEIELEDKFENM
VEVLAVKVTLPKGKGRNVVLDKSFAGAPTITKDGVSVAKEIELEDKFENM
IDILAEAVAVTLGPKGRNVVLEKKFGAPQIINDGVTIKIEIELEDHIENT
IDKLADAVGLTLGPRGRNVVLD.EFGSPKVVNEGVTIARAIELPDPMENA
VEKLANAVGVTLPGRGRNVVLD.EYGNPKVVNDGVTIARAIELPDAMENA
IDKLADAVGLTLGPRGRNVVLD.EFGSPKVVNDGVTIARAIELPDAMENA
VNKLADLVGVTLPKGKGRNVVLESKYGSPRIVNDGVTVAREVELEDPVENI

V--LADAV-VTLGPKGRNVVLE-FG-P-I--DGV-IA--IEL-D--EN-
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
150

GAKLVQDVANNTNEEAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAKLLQEVASKTNYAAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAELVKEVAKKTDVAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAELVKEVAKKTDVAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAQMVKEVASKANDAAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAQMVKEVASKANDAAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GVALIRQAASKTNDAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAALIREVASKTNDAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAALIREVASKTNDAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAALIREVASKTNDAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAKLVRQAAAKTNDAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GAKLVRQAAAKTNDAGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
GA-LV-EVA-KTND-AGDGTATVLAQALVKEGLRNVAAGANPLGLKRG
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
200

```



M1 VMLAVDAVIAELKKQSKPVTTPPEEIAQVATISANGDKEIGNIISDAMKKV  
M2 SQVAVEKVIEFLSANKKEITTSEEIAQVATISANGDSHVGLLASAMEKV  
P1 IEKAVDKVTETLLKDAKEVETKEQIAATAAISA.GDQSIGDLIAEAMDKV  
P2 IEKAVEKVTETLLKGAKEVETKEQIAATAAISA.GDQSIGDLIAEAMDKV  
P3 IDKAVTAAVEELKALSVPSCSDSKAIAQVGTISANSDETGVGLIAEAMDKV  
P4 IDKAVTAAVAELKKISKPCDKQKAIQVGTISANSDKSIGDIIAEAMEKV  
P5 IDKATNFLVEQIKSHARPVEDSKSIAQVGAI SAGNDFEVGQMIADAMDKV  
C1 IDKTVOGLIEELEKKARLLKVAMTLKLLLPFLERMSSLGRMIADAIDKV  
C2 IDKTVOGLIEELERKARPVKGSGDIKAVASISAGNDELIGAMIADAIDKV  
C3 IDKTVOALIEELEKRARPVKGGSDIKAVATISAGNDELVGTMIAADAIDKV  
C4 IEKTAKALVAELKKMSKEVEDS.ELADVAASAGNNAEIGSMIAEAMSRV

CON IDK-V---IE-L-----V-----IA-V--ISA--D--IG-MIA-AMDKV  
-----+-----+-----+-----+-----+ 250

GRKGVITVKDGKTLNDELEIIEGMKFDRGYISPYFINTSKGQKCEFQDAY  
GKEGVITIREGRTLEDELEVTEGMRFDGRFISKYINITDPKSDKVEFEKPL  
GNEGVITVEESNTFGLQLELTEGMRFDKGYISGYFVTD AERQEAVLEEPY  
GNEGVITVEESNTFGLQLELTEGMRFDKGYISGYFVTDPERQEAVLEDPY  
GKEGVITVEDGTGLQDELVDVEGMQFDRGYLSPYFINKPETGAVELESFF  
GKEGVITVEDGSGLENALVVEGMQFDRGYLSPYFINNQNM SAELENPF  
GKEGVISLEEGKSMTELEVTEGMRFDKGYISPYFATDTERMEAVFDEFF  
GPDGVLSIDSSSSFETTVEVEEGMEIDRGYISPOFVTNPEKLI CEFEKAR  
GPDGVLSIESSSSFETTVDVEEGMEIDRGYISPOFVTNLEKSIVEFENAR  
GPDGVLSIESSSSFETTVEVEEGMEIDRGYISPOFVTNPEKLLVEFENAR  
GRKGVVTLEEGKSAENALYVVEGMQFDRGYVSPYFVT DSEKMSVEFDNCK

G-EGVI-VEE---F---LEV-EGM-FDRGYISPYFVT-E-----EFE---  
-----+-----+-----+-----+-----+ 300

VLLSEKKISSIQSIVPALEIANAHRKPLVIAEDVDGEALSTLVNRLKV  
LLLSEKKISSIQDILPALEISNQSRRLIIAEDVDGEALAAACILNNLRG  
ILLVSSKVSTVKDLLPILLEKVIQAGKSLIIAEDVEGEALSTLVNKRIG  
ILLVSSKVSTVKDLLPILLEKVIQAGKPLIIAEDVEGEALSTLVNKRIG  
ILLADKKISNIREMLPVLEAVAKAGKPLIIAEDVEGEALATAVVNTIRG  
ILLVDKKISNIRELIPLLENVAKSGRPLLVIAEDIEGEALATLVNNIRG  
ILITDKKIGLVQDLVPVLEQVARAGRPLVIAEDIEKEALATLVNNRLRG  
VLVTDQKITAIKDIIPILLEKTTQLRAPLLIIAEDVTGEALASLVNKMGRG  
VLITDQKITSIKEIIPILLEQTTQLRCPLFIVAEDITGEALATLVNKLGRG  
VLITDQKITAIKDIIPILEKTTQLRAPLLIIAEDVTGEALATLVNKLGRG  
LLLVDKKITNARDLVGVLEDAIRGGYPILIIAEDIEQEALATLVNKLGRG

IL--D-KI--I-D--P-LE-----PLIIAEDVEGEALATLVN--RG  
-----+-----+-----+-----+-----+ 350



M1 GLQVVAVKAPGFGDNRKNQLKDMAIATGGAVFGEEGLTLNLEDVQPHDLG  
M2 QVKVCAVKAPNFGDNRKNTIGDIAVLTGGTVFTEE.LDLKPEQCTIENLG  
P1 TFKSVAVKAPGFGDRRKAMLQDMAILTGAQVISEE.VGLTLENTDLSLLG  
P2 TFKSVAVKAPGFGDRRKAMLQDMAILTGGQVISEE.VGLTLENADLSLLG  
P3 IVKVAAVKAPGFGDRRKAMLQDIATLTGGTVISEE.IGMELEKATLEDLG  
P4 VVKVAAVKAPGFGDRRKAMLQDIAVLTGGKVISEE.VGLSLEAASLDDL  
P5 VLNVAAVKAPGFGDRRKAMLEDIAVLTGGQLI  
C1 ILNVAAIKVPFGGERRKALLQDIAILT.GAEFQASDLGLLVENTSVEQLG  
C2 IINVAAIKAPSFGERRKAVLQDIAIVTGAEYLAKE.LGLLVENATVDQLG  
C3 VLNVAAVKAPGFGGERRKAMLQDIAILTGPST.T.LDMGLLVENTTIDQLG  
C4 TLKIAALKAPGFGGERKSQYLDIAILTGTATVIREE.VGLSLDKAGKEVLG

CON ---V-AVKAPGFGDRRKA-L-DIAILTG-----E--GL--E-----LG  
-----+-----+-----+-----+-----+-----+ 400

KVGEVIVTKDDAMLLKKGKDRAQIEKRIQEIIEQLDVTTSEYEKEKLNER  
SCDSITVTKEDTVILNGSGPKEAIQERIEQIKGSIDITTSYEKEKLQER  
KARKVMTKDETTIVEGAGDTDAIAGRVAQIRTEIENSDDYDREKLQER  
KARKVVTKDETTIVEGAGDTDAIAGRVAQIRQEIENSDDYDREKLQER  
QAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDYDREKLQER  
SAKRVVTKDDTTIIDGSGDAGDIKNRVEQIRKEIENSSDYDREKLQER

IARKVTITKDSTTLIADAASKDELQARIAQLKRELAETDSVYDSEKLAER  
TARKITIHTTTTTIADAASKDEIQARVAQLKKELSETDSIYDSEKLAER  
IARKVTISKDSTTLIADAASKETLQARISQLKKELSETDSVYDSEKLAER  
HAAKVVLTKETSTIVGDGSTQDAVQKRVTOIKNLIQAEQDYKEKLNER

-A--V---KD-TT-I-----I--RV-Q-----S-YD-EKL-ER  
-----+-----+-----+-----+-----+ 450

LAKLSDGVAVLKVGGTSDVEVNEKKDRVTDALNATRAAVEEGIVLGGGCA  
LAKLGGVAVIRVGGASEVEVGEKKDRYDDALFATRAAVEEGILPGGGTA  
LAKLAGGVAVIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVT  
LAKLAGGVAVIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVT  
VAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVGAGGVA  
LAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVA

IAKLSSGGVAVIKVGAATETELEDRLRIEDAKNATFAAIEEGIVPGGGAA  
IAKLSSGGVAVIKVGAATETELEDRLRIEDAKNATFAAIEEGIVPGGGAA  
MPKLPGGVAVIKVGAATETELEDRLRIEDAKNATFAAIEEGIVPGGGAT  
IAKLSSGGVAVIQVGAQTETELKEKKLRVEDALNATKAAVEEGIVVGGGCT

-AKL-GGVAVIKVGA-TE-EL-E-K-RIEDA--AT-AAVEEGIV-GGG--  
-----+-----+-----+-----+-----+ 500

M1 LLRCIPALDSLTPA..NEDQKIGIEIIKRTLKIPAMTIARNAGVEGSLIV  
M2 LVKASRVLDEV..VVDNFDQKLGVDIIRKAITRPAKQIIENAGEEGSVII  
P1 LLQAAPALDKLKTGDEA.T..GANIVKVALEAPLKQIAFNNGMEPGVVA  
P2 LLQAAPTLDKLEGDEA.T..GANIVKVALEAPLKQIAFNNGLEPGVVA  
P3 LIRVASKLADLRG..QNEQNVGIVKVALRAMEAPLRQIVLNCGEPSVVA  
P4 LIRVLKSLDSV..EVENEDQRVGVEIARRAMAYPLSQIVKNTGVQAAVVA  
C1 LVHLSTVVPAINGE.DDADERLGAD...RHWYPTSLIAQNAGIEGEVVV  
C2 YVHLSTYVPAIKETIEDHDERLGADIIQKALQAPASLIANNAGVEGEVVI  
C3 LVHLSTVIPAIKEKLEDADERLGADIVQKAL.VARSLIAQNAGIGGEVVV  
C4 LVCLASKVDAIKAALDNDEEKVGADIVKRALSYPKLIARNAGVNGSVVS

CON L-----D--D--G--I--AL--P---I--N-G----VV-  
-----+-----+-----+-----+-----+ 550

EKIMQ.SSSEV..GYDAMAGDFVNMVEKGIIDPTKVVRTALLDAAGVASL  
GKLIDEYGDDFAKGYDASKSEYTDMLATGIIDPFKVVRSGLVDASGVASL  
EKVRNLSVGH...GLNAATGEYEDLLKAGVADPVKVTRSAQNAAASIAL  
EKVRNLPAGH...GLNAQTGVYEDLLAAGVADPVKVTRSAQNAAASIAL  
NTVKG.GDGNY..GYNAATEEYGNMIDMGILDPKVTSAQNAAASIAL  
DKVLNKHKDVNY..GYNAATGEYGDMIEMGILDPKVTSAQNAAASIAL  
EKVKAREW.EI..GYNAMTDKYENLVEAGVIDPAKVTRCALQNS  
EKIKESEW.EM..GYNAMTDKYENLIESGVIDPAKVTRCALQNAASVSGM  
EKIMFSEW.EI..GYNAMTDTYENLLEAGVIDPAKVTRCALQNAASVAGM  
EKVL..SNDNVKFGYNAATGKYEVSMAAGIIDPTKVVRCCLEHAASVAKT

EKV-----GYNA-T--Y--L--G--DP-KV-R-AL--AA-VA-L  
-----+-----+-----+-----+-----+ 600

LTAEVVVTEIPKEEKD..PGMGAMGGMGGMGGMGGMF  
LATTEVAIVDAP..EPPAAAGAGGMPG..GMPGMPGMM  
FLTTEAVVADKPEKTAA..PASDPTGGMGGGMDF  
FLTTEAVVADKPEKEKASVPG.G..GDMGGGMDF  
MITTECMVTDLPKNDAAADLGAAGGMGGM.G..GMGGMM  
MITTECMVTEAPKKKEESMPGGGDMGGMGGMGGMGMM

VLTTQAIIVVEKPKPKKVAEPAEGQLSV  
VLTTQAIIVVDKPKPKAPTAAPPQGLMV  
FLMSDCVVVEIKEPEFVP

--T---V-D-P-----  
-----+-----+-----+-----+ 638

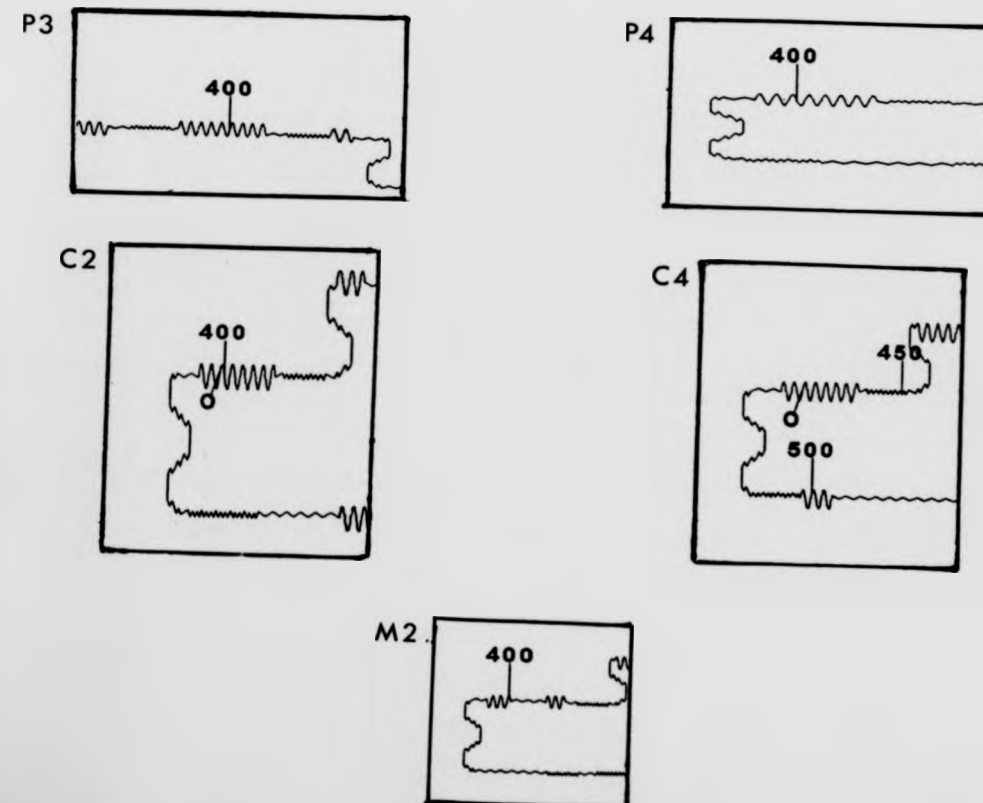
Figure 53. The reversed nucleotide binding site sequence present in the chaperonins and the secondary structure prediction

The chaperonin sequences are presented from the aminotermminus to the carboxyterminus (Panel A); M, mitochondria; P, prokaryotes; C, plastids. For detailed information on the different derived sequences see the legends to Figure 45, Figure 47 and Figure 50. The dinucleotide binding site fingerprint (Wierenga *et al.*, 1986) is written in the opposite direction i.e. from the carboxy to the aminotermminus (no 1 to 16). The various amino acids allowed at the different positions in the fingerprint are shown. Stars (\*) indicate amino acids that match the fingerprint and arrows (^) positions that do not match. In Panel B the secondary structure predictions of the reversed dinucleotide binding site of the *E.coli* (P3), the *C.burnetii* (P4), the *T.aestivum*  $\alpha$  subunit (C2), the *B.napus*  $\beta$  subunit (C4) and the *S.cerevisiae* (M2) chaperonins are shown. For detailed information on the secondary structure predictions see the legend to Figure 46.

A

		C	C			C	C										N
		M	M			M	M										C CQ
		V	V			V	V										M MT
		L	L			L	L										V VS
	E	I	I			I	I										L LH
	D	A	A			A	A			G		G	G	A	A	K	I IR
36	*	^	^	^	^	*	*	*	*	*	*	*	*	*	*	*	1
M1	VLKVGGTSDVEVNEKKDRVTDALNATRAAVEEGIVLGGGCALLRCIPALDS																
M2	VIRVGGASEVEVGEKKDRYDDALFATRAAVEEGILPGGGTALVKASRVLDE																
P1	VIKVGAAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALIRVASKLAD																
P2	VIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLLQAAPALDK																
P3	VIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLLQAAPTLDE																
P4	VIKVGAAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVALIRVLSLDS																
C1	VIKVGAAATELEDRKLRIEDAKNATFAAIEEGIVPGGGAALVHLSTVPA																
C2	VIKVGATTETELEDRLRIEDAKNATFAAIEEGIVPGGGAAYVHLSTYVPA																
C3	VIKVGAAATELEDRKLRIEDAKNATFAAIEEGIVPGGGATLVHLSTVIPA																
C4	VIQVGAQTETELKEKKLRVEDALNATKAAVEEGIVVGGGCTLVCLASKVDA																

B



has to be acknowledged that the  $\beta$ - $\alpha$ - $\beta$  structure is not predicted for this region in all chaperonins. However this does not exclude the possibility that this region is involved in ATP binding. Walker *et al.* (1982) identified two segments A and B, that are thought to be involved in ATP binding (Section 3.2.5.4). A variant of segment A, G X X X X X G K S/T (rather than G X X X X G K S/T) is present at position 254-263 in the human mitochondrial, the cyanobacterial and the plastid  $\beta$  chaperonin sequences, while segment B is not present in any of the sequences. Recently Chin *et al.* (1988) have reported a more variable consensus sequence for segment B i.e. R/K/H X<sub>5-8</sub>  $\Phi$  X  $\Phi$ <sub>2</sub> D/E, where X = various residues and  $\Phi$  = hydrophobic residues. This segment is present from residue 307 to 319 in the prokaryotic and plastid sequences but not in the mitochondrial chaperonins. However, in none of the chaperonins is the secondary structure predicted for this region, similar to the predictions for the segment B region in adenylate kinase, Rho, RecA protein or protease La, i.e.  $\alpha$ -helix turn  $\beta$ -sheet (Fig. 46; Fig. 48; Fig. 49; Fig. 51; Chin *et al.*, 1988). In addition a characteristic feature of segment B also includes an acidic residue or its amide counterpart approximately 4 residues from the amino acid lysine (Chin *et al.*, 1988). This acidic residue is not present in any of the chaperonin sequences, since position 303 is a hydrophobic amino acid (Fig. 52). It is concluded that no good match for segment B was found in any of the chaperonin sequences.

Taken together the chaperonin amino acid sequences contain highly conserved regions. One such region (Fig. 52, residue 440-505) is possibly involved in the binding of ATP since it contains the reverse sequence of a dinucleotide binding site and no other regions which have been reported to be involved in ATP binding, are present in the chaperonin sequences. Although the chaperonin sequences are very similar, secondary structure predictions indicate that the plastid chaperonins have a common structure distinct from that of the prokaryotic chaperonins, with the exception of the cyanobacterial chaperonin whose structure resemble that of the plastid chaperonins. This observation indicates that the cyanobacterial chaperonin is more homologous to the plastid chaperonins, which might be expected since chloroplasts are thought to

have originated from cyanobacteria. The mitochondrial chaperonins do not display a common secondary structure; the human mitochondrial chaperonin resembles a plastid chaperonin-like structure, while the yeast mitochondrial chaperonin structure is more similar to that of the prokaryotic chaperonins. More mitochondrial chaperonin sequences are needed to establish a possible common structure.

### 3.3.5 PROTEINS RELATED FUNCTIONALLY TO THE CHAPERONINS

Chaperonins are thought to recognize specific regions, possibly hydrophobic, in unfolded polypeptides, to influence their folding and assembly by a mechanism that is unknown but involves ATP (Ellis, van der Vies & Hemmingsen, 1989). Regions with protein and dinucleotide binding properties similar to those in the chaperonins may be present in other proteins with known function. An algorithm described by Lipman and Pearson (1985) provides the possibility of sensitive protein similarity searches between a newly sequenced protein and the available protein sequences present in the different databases. This algorithm, used in the computer program FASTP, first screens two sequences for similarity by looking for aligned identical amino acids and subsequently considers conserved amino acid replacements as well as identities, using an amino acid replaceability matrix (PAM 250 matrix) which is based on observed frequencies of amino acid replacements in evolutionarily related proteins. The result is an initial score, which is then optimized by allowing for insertions and deletions to give an optimized score. The program lists the sequences selected from the database in order or ranking similarities. The eleven chaperonin sequences were utilized to search the EMBL and PIR (NBRF) databases, using the algorithm described above, and the first fifty selected proteins of each search were used for further analysis, to determine possible biologically significant similarities. Since some of the chaperonin sequences are already present in the database, they were among the twenty-two sets of fifty selected proteins each, but were not included in the subsequent analysis.

The statistical significance of the initial and optimized scores is determined by the  $z$  value, where  $z = (\text{similarity score} - \text{mean score from the data base scan}) / (\text{standard deviation from the data base scan})$ . Lipman and Pearson (1985) reported the following guidelines to identify those similarities believed to be biologically significant, based on a large number of database searches. For the initial and optimized score;  $z > 3$ , possibly significant;  $z > 6$ , probably significant and  $z > 10$  significant. These values are similar to those suggested by other investigators (Doolittle, 1986). The above guidelines were used to identify statistical significant scores for the 1100 selected proteins (22 lists of 50 proteins each) which show similarity to one or more of the chaperonin sequences. A probable significant score can be interpreted in two ways; the compared amino acid sequences contain either a short region of highly identical amino acids or a long region with a low percentage similarity. The algorithm used does not distinguish between the two possibilities. However it is argued that if a particular protein sequence in the database shows similarity with several of the chaperonin sequences it is more likely to contain regions of possible biological significance, despite its statistically low significance score, than when it shows similarity with only one of the chaperonin sequences. Therefore those protein sequences that show similarity to more than two of the chaperonin sequences were identified with the aid of a specially designed algorithm (kindly made by J van der Vies); these are listed in Table 17 and Table 18.

The *groES* protein from *E.coli* shows similarity with all the eleven chaperonin sequences and the calculated  $z$  values for the initial scores vary from 13.9 for the *B.napus*  $\alpha$  chaperonin to 53 for the *E.coli* chaperonin, suggesting a similarity of biological significance. Further analysis of the similarity of the *groES* protein and the chaperonins is discussed later (Section 3.3.5.1). The calculated  $z$  values for all of the other protein sequences that show similarity to more than two chaperonin sequences are between 6 and 10, indicating probable biological significance. However further detailed analysis showed that those similarities are between large regions in the protein sequences which contain a low percentage of conserved and identical amino acids

Table 17. Proteins possibly related to the chaperonins selected from the EMBL database

The EMBL nucleotide data library was searched with the chaperonin amino acid sequences. For this purpose all the nucleotide sequences in the database were translated in six reading frames and similarities with the chaperonin sequences identified using the algorithm described by Lipman and Pearson (1985) which was used in a computer program called FASTP. The EMBL nucleotide sequence data library of 19th May 1989, containing 24,374 entries and 29,640,915 bases, was used. Selected proteins that show similarity to more than two chaperonin sequences were identified using a specially designed algorithm (kindly made by J. van der Vies) and the result is shown in this table. The abbreviations used for the chaperonins are similar to those described in the legend to Table 15. The z-values shown were calculated using the optimized score.

Table 17. Proteins possibly related to the chaperonins selected from the EMBL database

Sequence	cg. protein	number*	M1	M2	P1	P2	P3	P4	P5	C1	C2	C3	C4
groES	gene <i>E. coli</i>	ECGROES	35	31	36	38	53	42	35	26	30	30	29
glyceraldehyde-3-phosphate	dehydrogenase	M19694			5					5			5
myosin	heavy chain	OCMYBHR								10			11

\* Number in the EMBL data base.



Table 18. Possibly chaperonin related protein selected from the NBRF protein database

The chaperonin sequences were used to search the PIR (Protein Identification Resource) database, a registered mark of NBRF (National Biochemical Research Foundation), using the computer program FASTP. The PIR protein database of 31st March 1989 (Release 20.0) containing 5980 sequences and 1,681,392 amino acid residues, was used. Further analysis of selected sequences as described in the legend to Table 17 and in Section 3.3.5.

Table 18. Possibly chaperonin related proteins selected from the NBRF protein database

Protein	Number*	M1	M2	P1	P2	P3	P4	P5	C1	C2	C3	C4
DNA directed RNA polymerase	A27336				5	5	4		7	7	7	4
Myosin heavy chain	NWRBCA				7		7		9		12	9
Insulin receptor precursor	A05274			3	3	2	3					
t complex peptide 1b	B24059			16	16		13	11			15	
30S ribosomal protein-A	A27607	6	7			9	8		4			
apolipoprotein B100 precursor	LPHUB	6				6			8			7
cytochrome b	C30020			4					5	6	5	
fusion glycoprotein	VGNZA2			9	10			10				7
heat shock p70 (trypanosoma)	A25398	4									6	3
DNA K protein	IQECDK	6				4	5					

\*Number in the PIR (NBRF) protein database.

(<15%), with the exception of the 30S ribosomal A-protein from *E.coli*. This protein shows similarity to five of the eleven chaperonin sequences i.e. the human and yeast mitochondrial chaperonins, the chaperonins from *E.coli* and *C.burnetii*, and the chloroplast  $\alpha$  chaperonin from wheat (Table 18). The  $z$  value calculated from the optimized score is 9, in the comparison of the 30S ribosomal A-protein and the *E.coli* chaperonin. The short region of similarity is found at the amino terminus of both proteins. Of the first 23 amino acids a striking number of 21 residues are identical (91% identity; Fig. 50A). The similarity with the other chaperonin sequences, as well as the possible biological significance of this observation, is discussed later (Section 3.3.5.2).

The hsp70 heat shock proteins of *Trypanosoma brucei* (hsp70) and *E.coli* (*dnaK*) show sequence similarity to a number of different chaperonin sequences (Table 18). This similarity was investigated further since the hsp70 proteins are regarded as another class of the molecular chaperone family on the basis that the hsp70 proteins are not sequence-related to the chaperonins (Ellis, van der Vies & Hemmingsen, 1989). The statistical significances of the different comparisons were *T.brucei* hsp70/*S.cerevisiae* mitochondrial chaperonin,  $z = 6$ ; *T.brucei* hsp70/*B.napus*  $\alpha$  chaperonin,  $z = 6$ ; *T.brucei* hsp70/*B.napus*  $\beta$  chaperonin,  $z = 3$ ; *E.coli* *dnaK*/*S.cerevisiae* mitochondrial chaperonin,  $z = 6$ ; *E.coli* *dnaK*/*E.coli* chaperonin  $z = 4$  and *E.coli* *dnaK*/*C.burnetii* chaperonin,  $z = 5$ , indicating a possible biological significance; the  $z$  value were calculated from the optimized score. However these identified similarities are found between a number of different regions in the hsp70 proteins, and a number of different regions in the chaperonins, in which the percentage of identical amino acid is very low (<8%) (data not shown). These observations show that the hsp70 heat shock proteins are not related to the chaperonins by the criteria used to classify the molecular chaperones i.e. that the proteins within one class are sequence-related, since > 40% of the amino acids along the entire length of the chaperonin sequences are identical (Hemmingsen *et al.*, 1988). Although the percentage amino acid identity is low in regions that show similarity between the hsp70

proteins and the chaperonins, it may well be that some of these identical amino acids are important in the correct folding of particular domains. Since the hsp70 proteins and the chaperonins both belong to the family of molecular chaperones, the members of which are thought to share similar functions, it would not be unexpected to find similar conformations within these proteins. Information on the three-dimensional structure of the proteins is needed to establish whether the low similarities found by direct amino acid comparisons are of biological significance. A visual analysis of the twenty-two lists of selected proteins that are the result of the database searches leaves one with the impression that protein sequences which show similarity to either one or two of the chaperonin sequences, and which fall in the category of possibly biological significant similarities (i.e.  $z$  value between 3 and 10), are proteins whose function requires either ATP or another nucleotide, e.g. myosin, kinases, ATPases, dehydrogenases, and proteins that are involved in nucleic acid metabolism such as DNA and RNA polymerases.

#### 3.3.5.1 The *groES* protein from *Escherichia coli*.

The possible biological significance of the similarity between the *E.coli groES* protein and the chaperonins is of great interest since the *groEL* and the *groES* proteins are thought to interact *in vivo* (Tilly & Georgopoulos, 1982); mutations in either of the *groE* genes have been reported to abolish oligomeric protein and bacteriophage head assembly in *E.coli* (Tilly *et al.*, 1981; Goloubinoff *et al.*, 1989). Recently *groES*-like proteins have been found in *Mycobacterium bovis* (Yamaguchi *et al.*, 1988), *Mycobacterium tuberculosis* (Shinnick *et al.*, 1989), *Coxiella burnetii* (Vodkin & Williams, 1988) and *Anacystis nidulans* (Cozens and Walker, 1987). Amino acid comparison of the different *groES*-like proteins shows that they are all related and contain between 40-50% identical residues (Table 20 and Fig. 54), similar to the identities observed between the prokaryotic chaperonins (Table 15). The amino acid sequences of the *groES*-like proteins and the chaperonins were compared using an

Table 19. Percentage identical and conserved amino acids between the groES-like proteins and the chaperonin sequences

		groES-like protein				
Chaperonins		<i>M.bovis</i> (PS1)	<i>M.tubercu- losis</i> (PS2)	<i>E.coli</i> (PS3)	<i>C.brunetii</i> (PS4)	<i>A.nidulans</i> (PS5)
M1	mitochondrion ( <i>H.sapiens</i> )	36 (18)	37 (19)	36 (16)	40 (16)	43 (21)
M2	mitochondrion ( <i>S.cerevisiae</i> )	33 (15)	33 (22)	53 (16)	51 (16)	51 (20)
P1	<i>M.leprae</i>	36 (15)	37 (22)	44 (19)	43 (20)	38 (16)
P2	<i>M.tuberculosis</i>	36 (15)	37 (16)	44 (16)	43 (29)	47 (16)
P3	<i>E.coli</i>	38 (16)	38 (16)	41 (16)	48 (16)	40 (17)
P4	<i>C.brunetii</i>	38 (16)	38 (17)	42 (16)	44 (16)	44 (17)
P5	<i>A.nidulans</i>	35 (16)	36 (17)	43 (15)	50 (15)	41 (15)

		<i>groES</i> -like protein				
Chaperonins		<i>M.bovis</i> (PS1)	<i>M.tubercu- losis</i> (PS2)	<i>E.coli</i> (PS3)	<i>C.brunetii</i> (PS4)	<i>A.nidulans</i> (PS5)
C1	Alpha	32	33	43	46	49
	( <i>R.communis</i> )	(17)	(16)	(15)	(15)	(17)
C2	Alpha	34	32	41	46	36
	( <i>T.aestivum</i> )	(15)	(15)	(15)	(15)	(17)
C3	Alpha	35	34	44	46	54
	( <i>B.napus</i> )	(16)	(17)	(16)	(15)	(16)
C4	Beta	34	34	32	36	35
	( <i>B.napus</i> )	(16)	(16)	(16)	(18)	(17)

The numbers show percentages conserved and identical (in brackets) amino acids and were determined using the algorithms described by Needleman and Wunsch (1970) and Sellers (1974). For more information see also the legend to Table 15 and Table 16.

Figure 54. Amino acid sequence comparison of the *groES*-like proteins

The derived *groES*-like amino acids sequences from *M.bovis* (PS1; Yamaguchi *et al.*, 1988), *M.tuberculosis* (PS2; Shinnick *et al.*, 1989), *E.coli* (PS3; Hemmingsen *et al.*, 1988), *C.burnetii* (PS4; Vodkin & Williams, 1988) and *A.nidulans* (PS5; Cozens & Walker, 1987), were compared and a consensus sequence (CON) showing conserved amino acids is presented. Vertical lines show identities and dots are gaps introduced to optimize the alignment. The algorithm used was described by Needleman & Wunsch (1970).

```

PS1  MAKVNIKPLEDKILVQANEAEETTTASGLVIPDTAKEKPQEGTVVAVGP
PS2  MAKVNIKPLEDKILVQANEAEETTTASGLVIPDTAKEKPQEGTVVAVGP
PS3      MNIRPLHDRVIVKRKEVETKSAGGIVLTGSAAAKSTRGEVLAVGN
PS4      MKIRPLHDRVVVRRLEEERTSAGGIVIPDSAAEKPSRGEVISVGP
PS5  MAAVSLSVSTVTPLGDRVFKVAEAEKTAGGIILPDNAKEKPQVGEIVAVGP
CON  -----I-PL-D-V-V---E-E---A-G-V--D-A--K---G-V--VG-
      -----+-----+-----+-----+-----+-----+-----

```

53

```

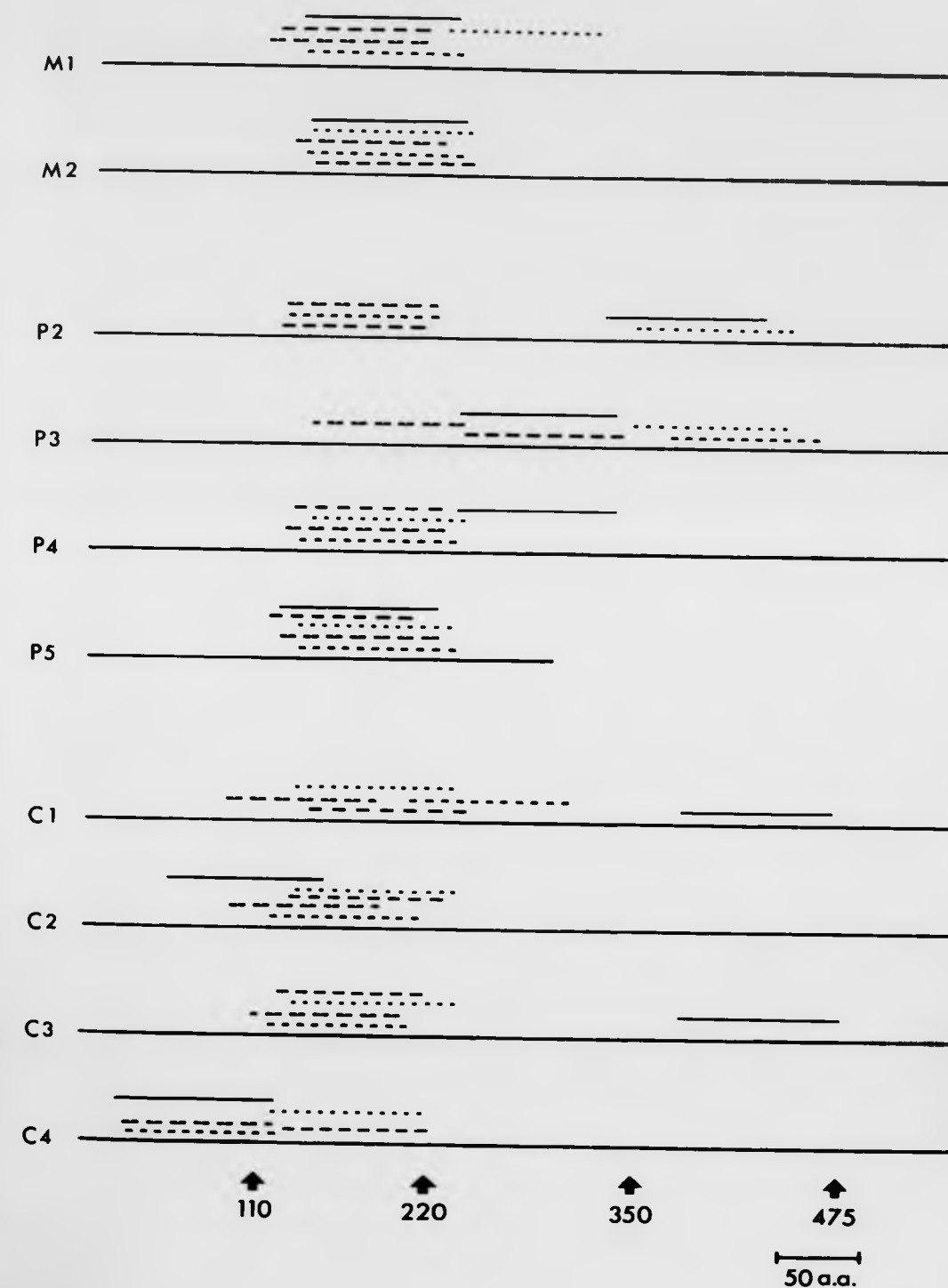
GRWDEGEKRIPLDVAEGDTVIYSK.YGGTEIKYNGEYLIILSARDVVGRRFK
GRWDEGEKRIPLDVAEGDTVIYSK.YGGTEIKYNGEYLIILSARDVLAVVSK
GRILENGEVK.PLDVKVGDIVIFNDGYGVKSEKIDNEEVLMSESDILAIVEA
GKPLDNGEVR.SLDVKVGDIILFGK.YAGTEVKLAGDEYIVMREDDIMGVIEK
GKRNDGSRQAP.EVKIGDKVLYSK.YAGTDIKLGNDYVLLSEKDILAVVA
G---E-G-----DV--GD-V-Y---Y-----K---EE---L---DI-----
      -----+-----+-----+-----+-----+-----+-----

```

106

Figure 55. Representation of the amino acid comparison of the *groES*-like proteins and the chaperonins

The approximate relative length of every chaperonin sequence is presented as a solid line. Amino acid residue 81 in Figure 52 is taken as the start of the mature chaperonin sequences which are consequently about 550 residues long. The position of the different shorter lines representing the *groES*-like proteins indicates where the amino acid similarity is found; *Mycobacterium bovis* (---); *Mycobacterium tuberculosis* (-----); *Coxiella burnetii* (.....); *Escherichia coli* (- - - -) and *Anacystis nidulans* (———). The numbers show approximate amino acid positions and are used to identify regions of similarity. The abbreviations for the chaperonin sequences are described in Figure 45, Figure 47 and Figure 50.





algorithm initially described by Needleman and Wunsch (1970) which scores for evolutionarily conserved amino-acid replacements in the compared sequences. The percentage similarity between the different chaperonins and the *groES*-like proteins varies from 32% to as high as 54%; the latter value is found with the *B.napus*  $\alpha$  chaperonin and the *A.nidulans* *groES*-like protein (Table 19). It is interesting to notice that a lower percentage similarity (41%) between the *E.coli* *groES* protein and the *E.coli* chaperonin is found than between the *E.coli* *groES* protein and the mitochondrial chaperonin from *S.cerevisiae* (53%). If the chaperonins and the *groES* proteins are evolutionarily related, they must have diverged a long time ago, as their amino acid identities are between 15% and 22% (Table 19). However the similarities were calculated to be biologically significant. In order to be able to identify particular regions in the chaperonin sequences that have similarity to one or more *groES*-like proteins, a representation of the different alignments was made (Fig. 55). A region at the aminoterminal end of the chaperonin sequences (see Fig. 55, between a.a. 100-220) shows similarity to the majority of the *groES*-like protein; some similarities are also observed in the region between residue 220 and 350, and between residues 350 and 475. The latter region contains the reversed dinucleotide binding site sequence, but similarity in this region was only observed with the cyanobacterial *groES*-like protein. To identify whether certain amino acids in the conserved chaperonin region are present in the *groES*-like protein sequences, more detailed analysis is required.

If the *groES*-like proteins contain identical amino acids which may be essential for their biological function, some structural similarity would be expected. However the secondary structure predictions show that the *groES* proteins have different conformations (Fig. 56). An attempt to identify secondary structures similar to the *groES* proteins in any of the chaperonins was not successful. It is not clear from the data presented here what the relationship between the *groES*-like proteins and the chaperonins is apart, from the fact that the similarities are identified as being biologically significant. It might be that the *groES*-like proteins originated as a result of duplication of a part of the chaperonins, possibly of the prokaryotic chaperonin at an

Table.20. Percentage identical and conserved amino acids of the *groES*-like proteins

Percentage identical amino acids

PS1	<i>M. bovis</i>	100				
PS2	<i>M.tuberculosis</i>	94	100			
PS3	<i>E.coli</i>	34	37	100		
PS4	<i>C.brunetii</i>	46	46	48	100	
PS5	<i>A.nidulans</i>	45	49	41	47	100
		PS1	PS2	PS3	PS4	PS5

Percentage conserved amino acids

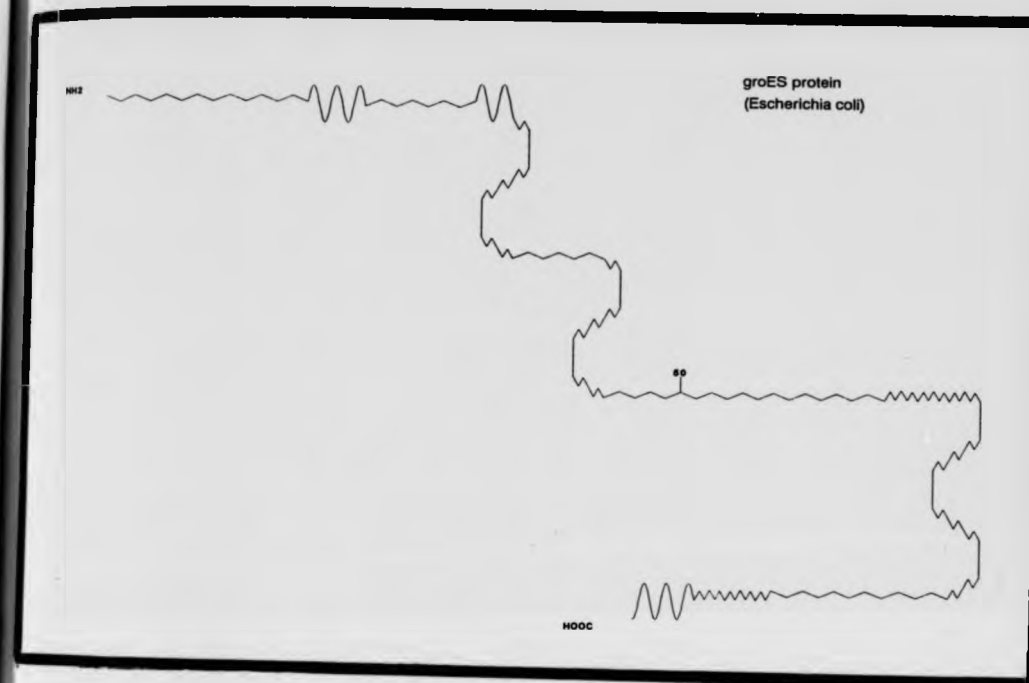
PS1	<i>M.bovis</i>	100				
PS2	<i>M.tuberculosis</i>	97	100			
PS3	<i>E.coli</i>	61	64	100		
PS4	<i>C.brunetti</i>	70	72	73	100	
PS5	<i>A.nidulans</i>	73	76	65	71	100
		PS1	PS2	PS3	PS4	PS5

The numbers show percentages and were determined using the algorithms described by Needleman and Wunsch (1970) and Sellers (1974). For detailed information see the legends to Table 15 and Table 16.

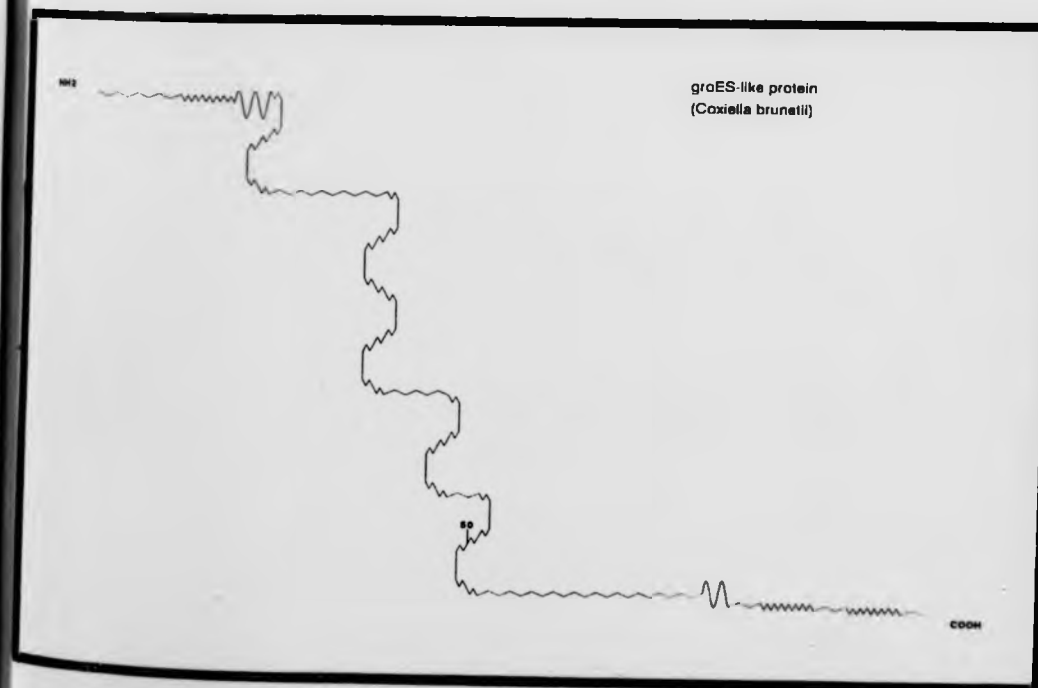
Figure 56. Secondary structure predictions of the *groES*-like proteins

The secondary structure predictions according to Chou & Fasman (1978) are presented as two-dimensional plots. Details of the different conformations is decribed in Figure 46.

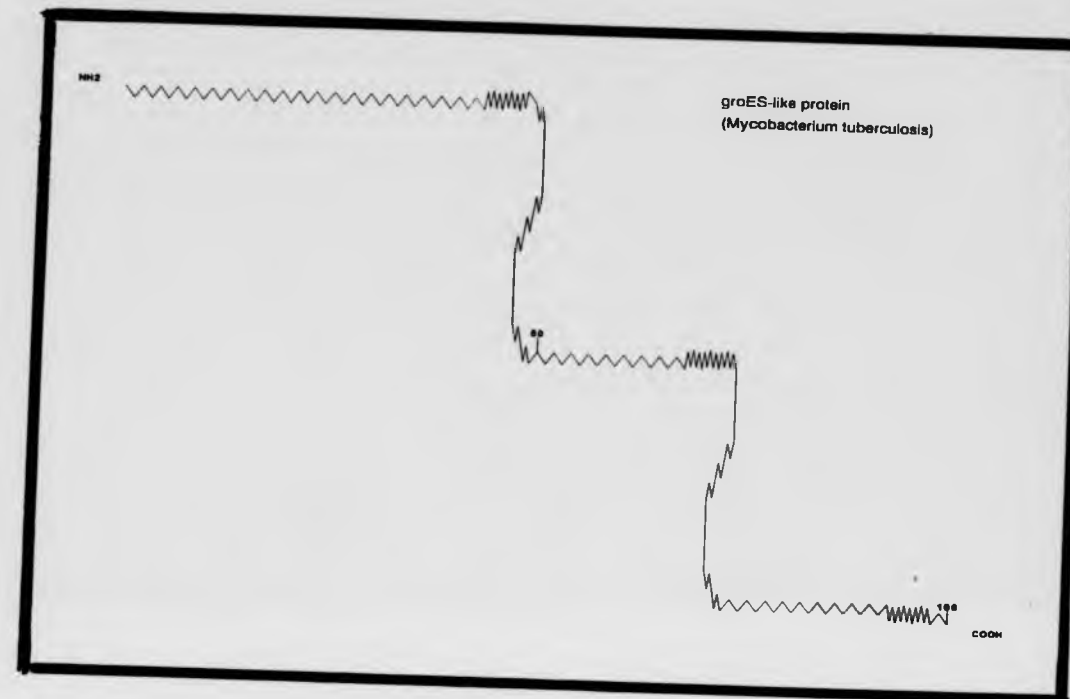
A



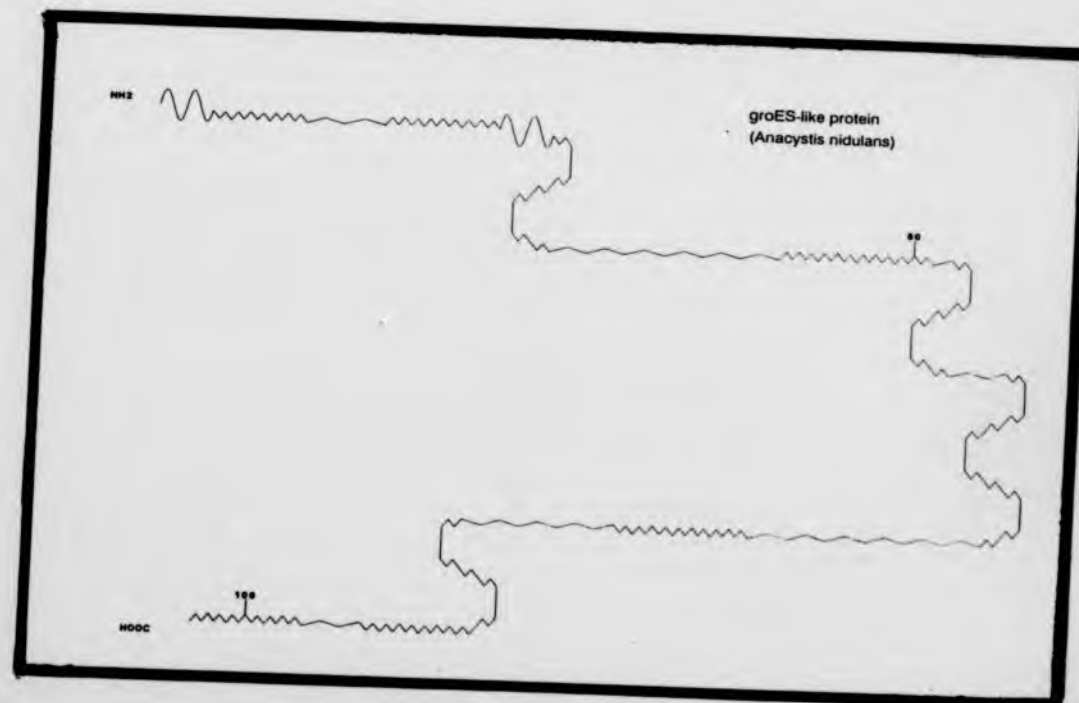
B



C



D



early stage in evolution. Recently, S.M. Hemmingsen (personal communication), who initially suggested to us a possible relationship between the *E.coli groES* and *groEL* proteins, has shown that a number of amino acids in the *groES*-like proteins are identical to amino acids present in a region at the aminoterminal end of the chaperonin sequences.

#### 3.3.5.2 The 30S ribosomal A-protein from *Escherichia coli*

The 30S ribosomal A-protein sequence and the *E.coli* chaperonin contain 91% identical amino acids in the first 23 residues of the aminoterminal of both proteins (Fig. 57A). The 30S ribosomal A-protein sequence is a determined aminoterminal sequence of a protein that is found associated with the 30S subunit of *E.coli* ribosomes and is not released by 1 M NH<sub>4</sub>Cl (Subramanian *et al.*, 1979). This A-protein was later identified as being identical to the *E.coli groEL* gene product (Neidhardt *et al.*, 1981) and it is thus not suprisingly that similarity with other chaperonin sequences is observed (Fig. 57B). It is not clear why the identity between the *E.coli* chaperonin and the A-protein is only 91%. Interestingly the amount of *groEL* protein associated with the ribosomes varies at different stages of the growth cycle, and the ratio of *groEL* protein to ribosomal protein increases fourfold between exponential and stationary growth. This observation can partly be explained by the fact that the ratio increases as the cells grow more slowly. Neidhardt *et al.* (1981) reported a twofold increase in this ratio in *E.coli* cells grown in glucose-rich medium compared to cells grown in acetate minimal medium. Since several biochemical changes take place when cells pass from active growth to a non-growing stationary phase, the increased amount of *groEL* protein associated with the ribosomes may implicate an involvement in the process of translation. Although there is no evidence that the *groEL* protein is directly involved in translation, ribosomes isolated from stationary phase *E.coli* cells are less active (80%) *in vitro* than ribosomes isolated from early and mid-log phase cells (Subramanian *et al.*, 1979).

Figure 57. Amino acid sequence comparison of the 30S ribosomal A-protein from *E.coli* and a number of chaperonins

Panel A shows the comparison of the *E.coli* chaperonin i.e. the *groEL* protein (P3) and the 30S ribosomal A protein (ribo). Stars indicate identities. In Panel B the 30S ribosomal A-protein from *E.coli* is compared with the mitochondrial chaperonins from *H.sapiens* (M1) and *S.cerevisiae* (M2), the *T.aestivum*  $\alpha$  chaperonin (C2) and the prokaryotic chaperonins from *E.coli* (P3) and *C.burnetii* (P4). Vertical lines indicate identical amino acids residues; stars (\*) show identical amino acid residues; asterisks (\*) show 80% amino acid identity; squares ( $\square$ ) show small/hydrophobic amino acids and the halved circles ( $\ominus$ ) show 80% acidic amino acid identity.

A

P3	AAKDVKFGNDARVKMLRGVNVLA
Ribo.	AAKDVKFGNDARVKMLEGVNRLA

B

	$\ominus$	$\ominus$																					
	*	*	*	$\square$	*	*	*	*	*	*													
M1	Y	A	K	D	V	K	F	G	A	D	A	R	A	L	M	L	Q	G	V	D	L	L	A
M2	S	H	K	D	L	K	F	G	V	E	G	R	A	S	L	L	K	G	V	E	T	L	A
C2	D	A	K	E	I	A	F	D	Q	K	S	R	A	A	L	Q	A	G	V	E	V	L	A
P4	A	A	K	V	L	K	F	S	H	E	V	L	H	A	M	S	R	G	V	E	V	L	A
P3	A	A	K	D	V	K	F	G	N	D	A	R	V	K	M	L	R	G	V	N	V	L	A
Ribo.	A	A	K	D	V	K	F	G	N	D	A	R	V	K	M	L	E	G	V	N	R	L	A



### 3.3.6 IDENTIFICATION OF THE REVERSED DINUCLEOTIDE BINDING SITE SEQUENCE IN OTHER PROTEINS

The reversed nucleotide binding site sequence is suggested to be involved in the binding of ATP (see Section 3.2.5.4). However no examples are known in which such a reversed sequence has been shown to bind a nucleotide molecule. It is unlikely for this reverse sequence to have been identified in other proteins since normally data base searches are performed with the protein sequence running in one direction only. Another complication in identifying similar reversed sequences is the variation which is allowed in the 'fingerprint' sequence e.g. the length of the loop (Wieringa *et al.*, 1986; Fig. 27). To preserve the variation in the different chaperonin reversed dinucleotide binding site sequences a similarity matrix was created that reflects the features of these sequences (Dix & Wootton, 1988). This matrix was subsequently used to search the OWL protein database for sequences that contain similar patterns (Bleasby *et al.*, 1988). The OWL database (version 1.1) was created by A.J. Bleasby (30th May 1988) and includes all the sequences in the NBRF (PIR) release 16.0, Genbank 54, Brookhaven, Oct. 1987, NEWAT86, SWISSPROT6.0 and JIPID adding up to a total of 10 777 entries and 2 694 603 amino acid residues. The sequences, selected from the OWL database, are listed in order of their degree of similarity (score) to the reversed dinucleotide binding site motif. The first fifty sequences were visually analyzed to identify a reversed dinucleotide binding site sequence, and this sequence was found in two proteins, the  $\text{Ca}^{2+}$ - $\text{Mg}^{2+}$ -dependent ATPase ( $\text{Ca}^{2+}$ -ATPase) from rabbit muscle sarcoplasmic reticulum, and the hormone-inducible proteins *Eip28* and *Eip29* (*Eip28/29*) from *Drosophila melanogaster* (Fig. 58C).

The motif in the  $\text{Ca}^{2+}$ -ATPase sequence is 34 amino acids in length, and matches 10 out of the 11 amino acid residues in the fingerprint only when one extra amino acid is allowed between the second and the third glycine. The motif present in the *Eip28/29* proteins also contains 9 out of the 11 essential residues represented by the



Figure 58. A reversed dinucleotide binding site sequence in the (Ca<sup>2+</sup>-Mg<sup>2+</sup>)-dependent ATPase and the ecdysome-inducible *Eip 28/29* proteins

All protein sequences are presented from the amino (left) to the carboxyterminus (right). In Panel A the reported dinucleotide binding site fingerprint (Wierenga *et al.*, 1986) is shown in the enzyme glyceraldehyde 3-phosphate dehydrogenase (GPDH). Panel B shows the reversed dinucleotide binding site fingerprint (read from right to left, 36 <--- 1). The abbreviations used for the chaperonins are explained in the legend to Figure 50. In both panels the allowed variation in the fingerprint motif is shown. Panel C shows two proteins, the Ca<sup>2+</sup>-Mg<sup>2+</sup>-dependent ATPase (CaATPase; MacLennan *et al.*, 1985) and the hormone-inducible protein *Eip 28/29* (Cherbas *et al.*, 1986). The dinucleotide binding site sequence should be read from right to left, 34 or 31 <-----1. Asterisks ( ) show amino acids that match the fingerprint whereas arrows ( ) represent mismatches. Notice that the length of the loop structures (residues 18 <---> 28 in Panel A is variable (see also Figure 27).

A

```

N
QC C          C C          C C
TM M          M M          M M
SV V          V V          V V
HL L          L L          L L
RI I          I I          I I E
KA A G G G    A A          A A D
1.....30
*****
GPDH SKIGIDGFGRIGRLVLRALSCGAQVVAVNDPFAILEYMV

```

B

```

C C          C C          C C Q
M M          M M          M MT
V V          V V          V VS
L L          L L          L LH
E I I        I I          I IR
D A A        A A          G G A AK
36.....1
* ^ ^      * *      * * * *
M1 VLKVGGS DVEVNEKKDRVTDALNATRAAVEEGIVLGGGCALLRCIPALDS
M2 VIRVGGASEVEVGEKKDRYDDALFATRAAVEEGILPGGGTALVKASRVLDE

P1 VIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALIRVASKLAD
P2 VIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLQAAAPALDK
P3 VIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLQAAPTLDE
P4 VIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVALIRVLKSLDS

C1 VIKVGAATETELEDRLRIEDAKNATFAAIEEGIVPGGGAALVHLSTVPA
C2 VIKVGATTETELEDRLRIEDAKNATFAAIEEGIVPGGGAAYVHLSTVPA
C3 VIKVGAATETELEDRLRIEDAKNATFAAIEEGIVPGGGATLVHLSTVIPA
C4 VIQVGAQTETELKEKKLRVEDALNATKAAVEEGIVVGGGCTLVCLASKVDA

```

C

```

34.....1
* * *      * ^ * * * *
CaATPase EFLQSFDEITAMTGDGVNDAPALKKAEIGIAMGSGTAVAKTASEV

31.....1
* ^ ^      * * * * *
Eip28/29 GIEQFKAEAETTGSDAHPAAVLLLPRGAERGPGSLLLTWPNLHRR

```

fingerprint sequence (Fig. 58C). It has been reported that some flexibility is allowed as far as the length of the loop structure of the nucleotide binding fingerprint is concerned (Wieringa *et al.*, 1986). An example of an enzyme containing a sequence that matches the fingerprint when read in the standard way from the amino to the carboxyterminus, is glyceraldehyde 3 phosphate dehydrogenase (GPDH) (Fig. 58A; Davidson *et al.*, 1967). The region between the small/hydrophobic amino acid preceding the loop structure and the small/hydrophobic amino acid following the loop structure (i.e. positions 18 and 28 in the original fingerprint respectively) is 8 amino acids in length rather than the 10 residues in the fingerprint. This sort of variation is allowed in nucleotide binding sites. In the  $\text{Ca}^{2+}$ -ATPase this region is 10 amino acids in length while in the *Eip28/29* proteins only 7 residues are found (Fig. 58C). The nucleotide binding site in GPDH is of particular interest since the purified oligomeric enzyme dissociates in the presence of ATP (Stancel & Deal, 1968). This ATP-dependent dissociation is also observed with the chaperonins. The role of ATP in the chaperonin function is discussed later in this thesis.

The *Eip28* and *Eip29* proteins are members of a family of polypeptides originally identified by their hormone-inducibility in *D.melanogaster* cell lines (Savakis *et al.*, 1980). The steroid hormone ecdysone causes proliferative arrest and morphological and enzymatic differentiation in cell lines that have been established from *D.melanogaster* embryos. The hormone response consists of enhanced synthesis of a small group of four polypeptides with apparent *M<sub>r</sub>*s of 28 000 (*Eip28*), 29 000 (*Eip29*), 35 000 and 40 000. The *Eip28* and *Eip29* proteins are encoded by a unique gene (*Eip28/29*) that gives rise to two distinct mRNA products by differential splicing (Cherbas *et al.*, 1986). The amount of *Eip28* and *Eip29* in untreated cells represents approximately 0.1% of the total cellular protein but is rapidly increased to a maximum of 1 to 2% upon treatment of the cells with ecdysone. The function of the *Eip28* and *Eip29* proteins is not known but the response noted above suggests that they may be stress proteins, like the chaperonins and other members of the molecular chaperone family (Hemmingsen *et al.*, 1988).

The  $\text{Ca}^{2+}$ -ATPase is a eukaryotic integral membrane ion-transporting enzyme, that directly couples the hydrolysis of ATP to the transport of ions across the membrane of the sarcoplasmic reticulum. Analysis of the deduced amino acid sequence showed the presence of probably six, and possibly eight, membrane spanning regions, and four cytoplasmic domains (MacLennan *et al.*, 1985). The cytoplasmic domains were identified between the residues 132-180, 330-505, 505-680 and 680-740. The amino acid sequence does not contain the typical segment A and segment B regions known to be involved in the binding of ATP in other proteins (Walker *et al.*, 1982). The reversed dinucleotide binding site sequence is present between residues 694 and 727 (Fig. 58C), and this region is part of an earlier identified cytoplasmic domain.

It is concluded that other proteins contain the reversed dinucleotide binding site sequence firstly identified in the wheat  $\alpha$  chaperonin; one such protein is the  $\text{Ca}^{2+}$ - $\text{Mg}^{2+}$ -dependent ATPase. This enzyme does not contain the ATP-binding sequences of Walker *et al* (1982) and must therefore have a different way of generating an ATP-binding site; this makes the identified reversed dinucleotide binding site sequence a candidate to be involved in the binding of ATP. The reversed sequence is found in only a small number of proteins which might indicate that identification of this flexible reversed fingerprint motif requires more advanced algorithms. However it is attractive to speculate that the reversed sequence is part of a special mechanism by which the chaperonins, and possibly a small number of other proteins, function.

### 3.3.7 DISCUSSION

This chapter describes the discovery of a new class of proteins which we have named the chaperonins. The proteins of this highly conserved class are structurally and immunologically related and ubiquitous in their occurrence in plastids, mitochondria and bacteria. The chaperonins comprise a class of the larger family of molecular chaperones as their known function, i.e. to assist the correct folding and assembly of other polypeptides, meets the criteria suggested for molecular chaperones. The

molecular chaperone family was proposed by Ellis (1987), who suggested that the essential function of molecular chaperones is to prevent the formation of improper structures which might result from the transient exposure of hydrophobic or charged surfaces normally involved in domain interactions, either within or between polypeptide chains. Such exposures may occur during the operation of several cellular processes such as the synthesis of polypeptides, the refolding that occurs during their transport across membranes, the association of polypeptides made in one subcellular compartment with those made in another, the changes in protein-protein interactions during the normal functioning of some complexes and during recovery from stress such as heat shock (Hemmingsen *et al.*, 1988). The other criterion that distinguishes molecular chaperones is that they do not form part of the final structure whose assembly they have mediated. Other suggested members of the molecular chaperone family (Ellis, van der Vies & Alldrick, 1989) are nucleoplasmin (Laskey *et al.*, 1978), the *hsp70*-immunoglobulin heavy chain binding protein group (Pelham, 1986; 1988), trigger factor (Crooke & Wickner, 1987; Crooke *et al.*, 1988), the *secB* protein (Collier *et al.*, 1988) and the signal recognition particle (SRP; Walter *et al.*, 1981). Recently two molecular chaperones have been described that are covalently attached to the molecule whose assembly they control, viz the pro-sequence of subtilisin (Zhu *et al.*, 1989) and the ubiquitin sequence present at the aminoterminal of two ribosomal precursor proteins in yeast and other eukaryotes (Finley *et al.*, 1989). This discussion is focussed on the chaperonin class and their function in several processes in the cell such as oligomeric protein assembly, protein transport, functioning of complexes (e.g. in DNA replication) and protein folding.

#### 3.3.7.1 Chaperonins and oligomeric protein assembly

The best-studied example of oligomeric protein assembly that requires the activity of a chaperonin is the assembly of the bacteriophage lambda head structure. The first step in the assembly pathway of bacteriophage lambda is the formation of an

oligomeric structure, known as the 'preconnector' (Fig. 30A and 30C), which consists of 12 identical phage-encoded polypeptides (pB). The preconnector is the basic structure on which the head proteins (pC and pE) assemble (Fig. 30B). The correct assembly of the preconnector structure requires the function of the *E.coli* chaperonin i.e. the *groEL* protein, as well as the function of the *groES* protein, the second polypeptide encoded by the *E.coli* *groE* locus, and the phage-encoded *pNu3* protein. The *groEL* protein is thought to interact directly with the phage B polypeptides, as early work by Georgopoulos *et al.* (1973) showed that  $\lambda$  mutants which have regained the ability to grow on *E.coli* *groE* mutants, contain mutations in gene B. The observation that a mutation in the *groEL* gene suppresses some mutations in the *groES* gene (Tilly & Georgopoulos, 1982), suggests that the two *groE* proteins interact in a functional manner. The role of the *pNu3* phage protein in the assembly of the preconnector structure is unknown, but appears to be essential since in the absence of a functional *pNu3* protein, phage B polypeptides are found associated with the *groEL* protein in extracts, and fail to form the oligomeric preconnector structure (Kochan & Murialdo, 1983). Neither the *groE* proteins nor the *pNu3* protein form part of the final preconnector structure. Evidence presented here (Section 3.3.5.1) suggests that the *groEL* and the *groES* protein may be related as they show amino acid sequence similarity. This raises the question whether the *groES* protein should also be classified as a chaperonin. However there is no evidence that the *groES* protein interacts with any other protein apart from the *groEL* protein. The amino acid sequence of the *pNu3* protein is not related to either the *groEL* or the *groES* protein (data not shown).

At a later stage of prohead assembly the *pNu3* protein acts as a core or scaffolding, but does not form part of the final prohead structure (Fig. 30B; Ray & Murialdo, 1975). The *groEL* protein too is likely to be involved at a later stage of prohead assembly since it has been shown with the aid of a series of *E.coli* mutants that the *groEL* protein interacts with the major head protein pE (Fig. 30B; Georgopoulos *et al.*, 1973).

The normal function of the *groE* proteins in uninfected cells is not known. However the observation that many of the *groE* mutations affecting phage growth do not inhibit cell growth (Georgopoulos *et al.*, 1973; Sternberg, 1973) suggests that the *groE* proteins perform either a different function for host and phage, or display a similar function and only a certain minimum 'activity' is essential for survival. The latter is supported by the fact that all known *groE* mutants are leaky and by the inability to isolate *groE* deletion mutations. Recently Van Dijk *et al.* (1989) showed that overexpression of both the *groE* proteins suppresses certain temperature sensitive mutations in a number of genes encoding oligomeric enzyme complexes involved in the biosynthesis of the branched chain amino acids leucine, isoleucine and valine in *Salmonella typhimurium* and the biosynthesis of histidine<sup>5</sup> in *E. coli*, while no suppression was observed of mutations in the genes encoding a number of monomeric aminoacyl-tRNA synthetase enzymes. These observations suggest that the *groEL* and *groES* proteins probably interact with a number of different oligomeric enzyme polypeptide chains, presumably to assist their correct folding and assembly.

The assembly of the oligomeric enzyme Rubisco, which consists of two types of subunits, is another example of a process which is thought to require the activity of a chaperonin. Eight large and eight small subunits comprise the Rubisco enzyme in most cases and the binding of these subunits to the chloroplast chaperonin has been demonstrated (Barracough & Ellis, 1980; Section 3.4; Ellis & van der Vies, 1988; Gatenby *et al.*, 1988). The most convincing evidence that this binding is a step in the assembly pathway of Rubisco is the observation that the transfer of newly-synthesized Rubisco large subunits to the holoenzyme in stromal extracts of *Pisum sativum* is inhibited by the addition of antibodies against the chloroplast chaperonin (Milos & Roy, 1984; Cannon *et al.*, 1986).

In contrast to the Rubisco enzyme from plants, the purified Rubisco from cyanobacteria can be dissociated into its subunits and subsequently reassembled *in vitro* (Andrews & Ballment, 1983). However this does not exclude the possibility that the assembly of Rubisco *in vivo* requires the activity of a chaperonin because

dissociation *in vitro* produces stable octameric large subunits. This idea was recently confirmed by Goloubinof *et al.* (1989) who demonstrated that the assembly of cyanobacterial Rubisco subunits synthesized from cloned genes in *E.coli* depends upon both the endogenous *groEL* and *groES* proteins. The high degree of amino acid identity between the *groEL* and *groES* proteins from *A.nidulans* and *E.coli*, (78% and 41% respectively) may account for their complementarity in this process of oligomeric protein assembly (Table 15). The failure of the *E.coli* chaperonins to mediate the assembly of wheat Rubisco (Bradley *et al.*, 1986) appears to be the reflection of the relatively low percentage of amino acid identity between the chloroplast and bacterial chaperonin (46%) as well as possible need for a distinct chloroplastic *groES*-like protein. The relatively low percentage of amino acid identity between the *E.coli* *groE* proteins and their mycobacterial homologues (58% and 37%) for the *groES* and *groEL* proteins respectively may also explain the failure of the mycobacterial chaperonin (65 kDa common antigen) to complement *E.coli* *groE* mutants that are blocked in the assembly pathway of bacteriophage lambda (D.B. Young, personal communication).

### 3.3.7.2 Chaperonins and protein transport

Polypeptides that are transported across either bacterial membranes or membranes of organelles such as the endoplasmic reticulum, mitochondria and plastids do so post translationally in an unfolded conformation (Eilers & Schatz, 1988; Wickner, 1989). Chaperonins have been suggested to mediate the folding and unfolding of polypeptides on possibly both sides of these membranes (Hemmingsen *et al.*, 1988). This idea was recently supported by the observation that newly-synthesized polypeptides in extracts of *E.coli* are associated with the chaperonin, in a conformation suitable for import into reversed inner-membrane vesicles (Bochkareva *et al.*, 1988). Chaperonins may also be present in the periplasmic space between the inner- and outer bacterial membranes, where they either refold and/or assemble translocated polypeptides, or assist protein folding to ensure translocation across the outer



membrane. Preliminary results suggest that a protein immunologically related to the *groEL*-protein is probably present in the periplasmic fraction of *Erwinia caratovora* (data not shown).

The involvement of the *groES* protein in the unfolding and refolding of polypeptides is not clear. However the purified *E.coli* chaperonin is capable of binding unfolded precursor polypeptides in a conformation that is suitable for cytoplasmic membrane translocation *in vitro* without the aid of the *groES* protein (Lecker *et al.*, 1989). Although the *groES* protein may not be required *in vitro*, this does not exclude the possibility that this protein is involved *in vivo*. This idea is supported by the observation that a temperature-sensitive mutation in the *E.coli* *secA* gene is suppressed only by the intact *groE* operon (Van Dijk *et al.*, 1989). The *secA* gene encodes a cytoplasmic protein, with a predicted Mr of 101 902, which is required for protein export in *E.coli*. The *secA* mutation (an amino acid change at position 43 that changes Leu → Pro) severely disrupts the predicted secondary structure of the protein (Schmidt *et al.*, 1988). The function of the *secA* protein is not known.

The proposed requirement for chaperonins in the refolding and assembly of polypeptides that are transported across the membranes of organelles (Ellis, 1987; Hemmingsen *et al.*, 1988; McMullin and Hallberg, 1988) has recently been confirmed as studies of a yeast strain carrying a mutation in the chaperonin gene have shown that this protein is required for the assembly of mitochondrial oligomeric protein structure *in vivo* such as F1 ATPase and ornithine transcarbamylase (OTC) (Cheng *et al.*, 1989). The polypeptides are found associated with the chaperonin after import into mitochondria *in vitro* (J. Martin, personal communication). This observation is strikingly similar to our earlier observation that Rubisco small subunits, imported into isolated chloroplasts from *Pisum sativum*, are associated with the chloroplast chaperonin (Section 3.4.3; Ellis & van der Vies, 1988). Although a wide range of polypeptides such as chloramphenicol transferase (CAT), glutamine synthetase (GS) and the  $\beta$  subunit of the thylakoid ATP synthase (CF1- $\beta$ ) are found associated with the chaperonin after transport across the chloroplast envelope membranes, not all imported

polypeptides associate with the chaperonin e.g. ferredoxin (FD) and superoxide dismutase (SOD) (J. Lubben *et al.*, 1989). It is possible that the presequence may function as a signal that is required or enhances the binding of polypeptides to the chaperonin, and this suggestion is strengthened by the report that both the presequence and regions within the mature *E.coli* outer membrane protein A (ompA) participate in the recognition of the *E.coli* chaperonin *in vitro* (Lecker *et al.*, 1989). It is not known whether *groES*-like proteins are involved in the refolding of imported polypeptides in organelles. Attempts to identify *groES*-like proteins in chloroplasts and mitochondria from plants and yeast, using antibodies against the *E.coli groES* protein, have so far been unsuccessful (J. Martin and R.L. Hallberg, personal communication; data not shown). As in *E.coli*, chaperonins or other molecular chaperones may be required in the intermembrane space of chloroplast and mitochondrial envelope membranes and the thylakoid lumen in chloroplasts, but so far no examples have been reported.

### 3.3.7.3 Chaperonins and nucleic acid metabolism

The first step in the replication of chromosomal DNA in *E.coli* is the cooperative binding of 20 to 30 *dnaA* proteins to a specific region (*oriC*) in the double-stranded DNA molecule both *in vivo* (Tomizawa & Selzer, 1979) and *in vitro* (Fuller *et al.*, 1981; 1983; 1984; Fuller & Kornberg, 1983). Mutations in the *groEL* and *groES* genes interfere with DNA synthesis at high temperatures (Wada & Itikawa, 1984), while overproduction of both the *groEL* and *groES* protein suppresses a (Ts) mutation in the *dnaA* gene, called *dnaA46* (Fayet *et al.*, 1986; Jenkins *et al.*, 1986). This mutation has recently been identified as a single amino acid change (Ala → Val) in the ATP-binding domain of the protein (Hwang & Kaguni, 1988a), and this change will have an effect on the activation of the protein by ATP, which is required prior to its binding to double-stranded DNA (Sekimizu *et al.*, 1987). The observation that the activity of the purified *dnaA46* protein *in vitro* is restored when DNA replication is assayed in crude enzyme extracts rather than a purified enzyme system led Hwang &

Kaguni (1988b) to suggest the involvement of a stimulatory factor which they subsequently showed to be a protein. This unidentified protein may well be the *E.coli* chaperonin, and this would suggest a function that involves the modulation of either protein-protein, protein-DNA or protein ligand interactions required for DNA replication.

Another protein that is involved in DNA replication is the *ssb* protein which binds preferentially and cooperatively to single-stranded DNA (Chase & Williams, 1986), but interacts also with a variety of other proteins such as DNA polymerase II (Molineux & Geftner, 1974; 1975). The *ssb* protein has an apparent Mr of 17 000 and appears as a functional tetramer that dissociates into non-functional monomers depending on the cellular concentration of the protein. Certain mutations in the *ssb* gene are suppressed by a mutation in the *groEL* gene (Ruben *et al.*, 1988). One of these mutant proteins *ssb-1* contains a single amino acid replacement (His-55 → Tyr) in a region of the DNA-binding domain (Williams *et al.*, 1983). The tetrameric structure of the *ssb-1* mutant protein appears to be unstable and it dissociates into monomers as the protein concentration decreases (Williams *et al.*, 1984). At the normal cellular concentration of the wild-type protein the mutant protein will be mainly present in the monomeric non-functional form. An attractive explanation for the observed suppression of the *ssb-1* mutant gene, is that the mutant *groEL* protein interacts with the mutant *ssb-1* protein either to suppress dissociation of the tetramers or to reassociate the monomers into functional molecules.

The *groE* proteins are suggested to be involved in the process of cell division as overexpression of the *groES* protein suppresses a mutation in the *letA* gene present on the sex factor plasmid F (Miki *et al.*, 1988). The product of the *letA* gene controls the coupling between DNA replication of the complete *E.coli* chromosome and the replication of the F plasmid, both of which are an essential prerequisite for cell division.

The *groE* proteins are also involved in RNA metabolism as certain mutations in the *groEL* and *groES* genes were shown to interfere with the synthesis of RNA at high

temperatures (Wada & Itikawa, 1984). The *groEL* protein appears to be involved in the stability of RNA in *E.coli*, since an internal fragment of the *groEL* gene suppresses a temperature-sensitive mutation in (*ams*<sup>+</sup>), a gene that influences messenger RNA half-life (Chanda *et al.*, 1985). This internal fragment may contain all the information that is required for the chaperonin function. It is therefore interesting to notice that this functionally active fragment contains the identified reversed dinucleotide binding sequences (see Fig. 52, P3, amino acid 389-505; Fig. 53).

It seems plausible to suggest from all the available evidence in prokaryotes that the chaperonins may also mediate protein-protein interactions involved in DNA replication and mRNA stability in chloroplasts and mitochondria. The presence of sequences homologous to the bacterial *dnaA* have recently been identified in the chloroplast genome of *Chlamydomonas reinhardtii* (Opperman *et al.*, 1989) which suggests that chloroplast DNA replication may be mediated by a protein that is structurally and functionally similar to the bacterial protein. However at the moment there is no direct evidence that chaperonins are involved in the replication of organelle genomes, and this possibility deserves attention.

#### 3.3.7.4 Chaperonins and protein synthesis

In the process of protein synthesis the aminoterminal region of a polypeptide is made before the carboxyterminal region and this situation may result in the incorrect folding of the polypeptide chain. Thus chaperonins (or molecular chaperones) may be required during protein synthesis. There is no evidence so far that elongating polypeptide chains bind to chaperonins. However since the *E.coli* chaperonin is found associated with ribosomes the chaperonins may be able to guide protein synthesis in some way (Subramanian *et al.*, 1979). The amount of *groEL* protein associated with the ribosomes varies depending on the growth rate and the phase of the cells. Isolated ribosomes with an increased amount of *groEL* protein associated are less active *in vitro*

than ribosomes that contain low amounts of the chaperonin (Subramanian *et al.*, 1979; Neidhardt *et al.*, 1981).

#### 3.3.7.5 Chaperonins as stress proteins and immunogens

The word stress is poorly defined in biological terms but is best thought of as a hostile situation that causes a specific response of either a cell or organism. Many kinds of stress such as heating, chilling, wounding and infection either elicit or enhance the expression of specific genes. All organisms examined respond to heat by inducing the synthesis of a group of proteins called heat-shock proteins (reviewed by Lindquist & Graig, 1988). The rate of protein folding and assembly, and the optimal functioning of a great number of enzymically active protein complexes is often restricted to a narrow temperature range and increasing the temperature is most likely to affect these processes. An increased rate of protein folding will increase the chance of the formation of incorrect protein structures occurring, and an increased need for proteins like the chaperonins is thus not unexpected. Extremes of heat cause protein denaturation and aggregation and it has been argued that one role of stress proteins is to unscramble these aggregates (Pelham, 1986). The effect of heat shock on the synthesis of chaperonins varies in extent but invariably causes enhanced accumulation. The concentration of the plastid and mitochondrial chaperonins increases two to three-fold (McMullin & Hallberg, 1987; S.M. Hemmingsen, personal communication). The *groE* genes in *E.coli* are members of a heat shock regulon, of which the promoter controlling transcription is recognized by RNA polymerase carrying the heat shock sigma factor,  $\sigma_{32}$  (Cowing *et al.*, 1985). The synthesis of the *groEL* protein, which accounts for 1% of cellular protein synthesis in cells in steady state growth at 37°C, increases to 10% of total synthesis soon after the cells are shifted to a growth temperature of 46°C (Hemmingsen *et al.*, 1988).

Another type of stress is encountered by the pathogenic bacteria, which are thought to respond to their hostile environment provided by their hosts, by increasing

the synthesis of the chaperonin protein, and other members of the molecular chaperone family such as the hsp70 group. These stress proteins elicit a strong antibody and T-cell response in patients infected with *Mycobacterium leprae* and *Mycobacterium tuberculosis* (Young *et al.*, 1987; Young, 1988; Young *et al.*, 1988). The potential danger that lies in the existence of conserved proteins present in all organisms is the development of an autoimmune response. The immune response to the bacterial chaperonin has been implicated in the aetiology of the autoimmune disease rheumatoid arthritis (van Eden *et al.*, 1988; Lamb & Young, 1989).

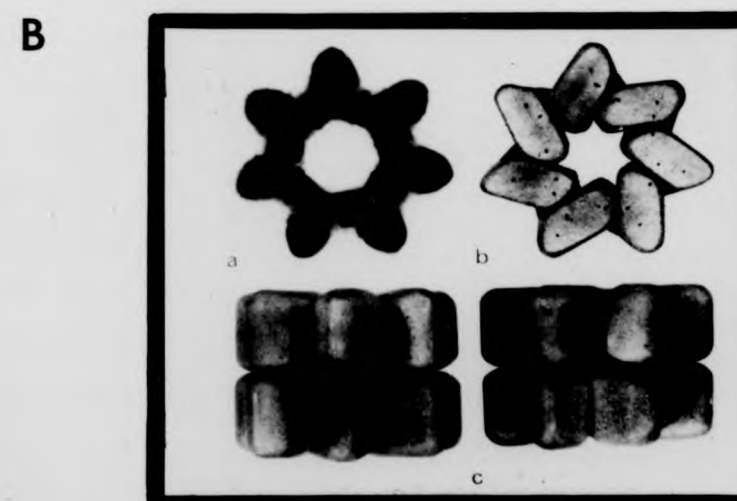
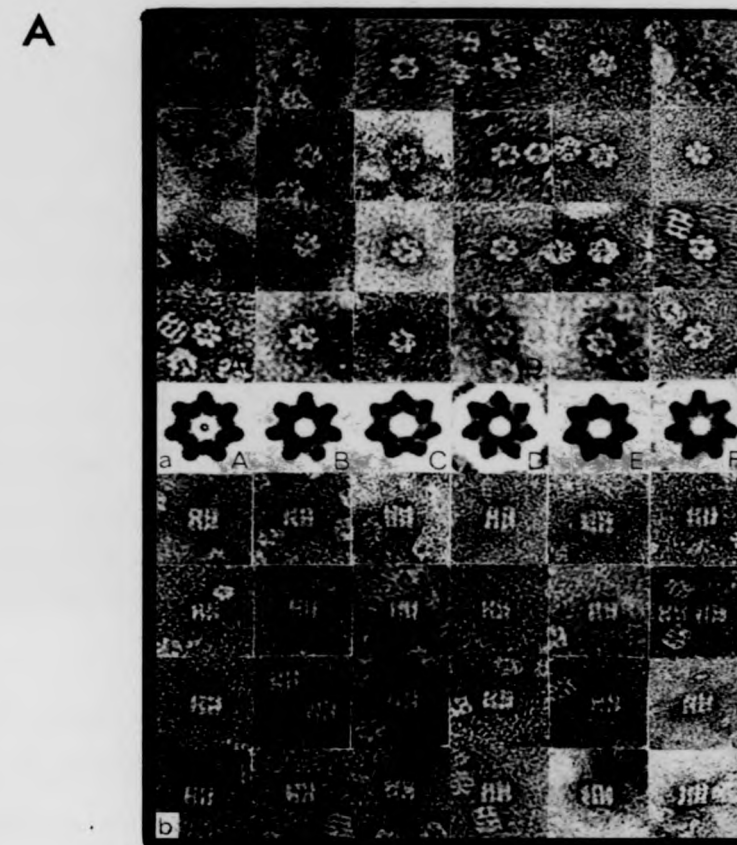
#### 3.3.7.6 Chaperonin ultrastructure and location

The chaperonins purified from *E.coli*, *P.sativum* chloroplasts, and from mitochondria of *T.thermophila*, *S.cerevisiae* and *Z.mays*, are all oligomeric protein complexes composed of 14 subunits that are arranged in two stacked rings of 7 subunits each (Fig. 59; Hendrix, 1979; Puskin, 1982; McMullin & Hallberg, 1988; Prasad & Hallberg, 1989). While the *E.coli* chaperonin, and probably the mitochondrial chaperonin, are homo oligomers, the chloroplast chaperonin is composed of two types of related subunits,  $\alpha$  and  $\beta$  (Table 15; Hemmingsen & Ellis, 1986; Ellis *et al.*, 1987). How the  $\alpha$  and  $\beta$  subunits are arranged in the chloroplast chaperonin oligomer is not known, but they are present in equal amounts in the purified protein (Musgrove & Ellis (1986).

Recent immuno-gold labelling studies of the chaperonin in the photosynthetic prokaryote *Chromatium vinosum* revealed that the protein is located in the cytoplasm and along the cell envelope (McFadden *et al.*, 1989). Although the distribution of labelled chaperonin was reported to vary considerably from cell to cell, the greatest concentration of protein was always found along the envelope as were the Rubisco large and small subunits. Studies to localize the chaperonin in plant cells are in progress (B. Andersson, personal communication), since it is not known whether this protein occurs in other locations than the stroma of plastids.

Figure 59. Structure of the chaperonin from *Pisum sativum*

The quaternary structure of the protein viewed through the electron microscope is shown. Magnification,  $\times 500\,000$ . Panel Aa; round particles with a rotational symmetry axis of 7, measuring  $12.3 \pm 0.8$  nm across the diameter. Original micrographs of the particles are marked A, B, C, D, E, and F. Panel Ab; side projection of the particles showing four distinct stripes. Panel B: Model images of the protein molecule; (a) result image of the photo-superposition of a number of particles; (b) model image of the molecule downward view along the seven-fold axis; side views along the two-fold axis in the two counter-wise directions. Photographs were taken from the paper by Pushkin *et al.* (1982), who purified the protein from *Pisum sativum* leaves but did not establish either its chloroplast location or its involvement in Rubisco assembly.





### 3.3.7.7 Chaperonins: mechanism of action

The mechanism by which the chaperonins and other molecular chaperones function is currently not understood. It has been proposed that during protein folding and assembly processes there is a certain probability that non-specific interactions will occur that may lead to the formation of incorrect, non-functional structures (Ellis & Hemmingsen, 1989). It is suggested that chaperones bind to transient exposed protein surfaces in order to prevent the formation of incorrect structures. The affinity of the molecular chaperone for such surfaces is dependent on their concentration and accessibility, i.e. the unfolded state of the polypeptide (Bochkareva *et al.*, 1988; Lecker *et al.*, 1989). The *E.coli* and *P.sativum* chaperonin show weak ATPase activities (Hendrix, 1979; Puskin *et al.*, 1982; Chaudhari *et al.*, 1987) and their oligomeric structures are both affected by the presence of MgATP. The dissociation of the chloroplast chaperonin oligomer in the presence of the MgATP is enhanced at lower temperatures (Hemmingsen & Ellis, 1986) and is reversible in stromal extracts when the concentration of ATP is lowered by allowing protein synthesis to occur. However, this reversal of dissociation is not observed if the ATP is removed by dialysis, suggesting that additional dialysable factors may be required for reversal (Musgrove & Ellis, 1986). Both the purified *P.sativum* and *E.coli* chaperonin dissociate in the presence of MgATP (Hemmingsen & Ellis, 1986; A.A. Gatenby, personal communication). The position of the equilibrium of this reversible dissociation *in vivo* is unknown, since the concentration of ATP inside the chloroplast and the *E.coli* cell is estimated to be in the range of 1-5 mM (Hampp *et al.*, 1982; Neuhaard & Nygaard, 1987), which is expected to favour dissociation, whereas the concentration of the chaperonin *in vivo* is estimated to be about 10 mg/ml and increases to 30-100 mg/ml upon heat shock. According to the law of mass action increasing the concentration of the chaperonin strongly favours association of the subunits. Thus results obtained *in vitro* cannot be extrapolated directly to the *in vivo* situation. In cells of *Tetrahymena*

*thermophila* the concentration of chaperonin increases two to threefold, upon heat shock as a result of increased protein synthesis. However the cellular concentration of ATP immediately decreases to 50% and this concentration is maintained at the heat shock temperature (Findly *et al.*, 1983). The decrease in intracellular ATP, combined with the increase in the chaperonin concentration should favour association. Because of the increased accumulation of the chaperonins under heat-stress it seems likely that the oligomeric form of the chaperonin is required to bind other partly unfolded polypeptides. The purified oligomeric *groEL* protein is capable of binding unfolded polypeptides in the absence of ATP *in vitro* (Lecker *et al.*, 1989).

Chandrasekhar *et al* (1986) reported that the purified *E.coli groEL* and *groES* proteins associate in the presence of MgATP, which may reflect the requirement for the destabilization of the interactions of the *groEL* monomers themselves. However their presented data also show that under these conditions a proportion of the *groEL* oligomers dissociate. This double effect of ATP may suggest a requirement of ATP for both the chaperonins activity *per se*, and the regulation of this activity. A precedent for an enzyme which activity is regulated by means of reversible ATP-dependent dissociation is the enzyme glyceraldehyde 3-phosphate dehydrogenase (Stancel & Deal, Jr., 1969). The dissociation of the tetrameric enzyme in the presence of 1 mM ATP into nonfunctional monomers is considered to be a means of regulating the activity of this enzyme in the glycolysis pathway.

How the chaperonins bind ATP is not known, but the reversed dinucleotide binding site sequence initially identified in the wheat  $\alpha$  chaperonin is present in all chaperonins examined (Section 3.3.4.5). The observation that this site is uniform in all chaperonins, and the absence of any other sequences known to be involved in nucleotide binding, suggests strongly that this site is involved in the binding of ATP. A search through the protein data bank revealed that other proteins such as  $\text{Ca}^{2+}$ -ATPase and the ecdysone inducible protein *Ei28/29* also contain this reversed nucleotide binding site sequence (Section 3.3.6). It is not known however whether these regions are actually involved in nucleotide binding. Further investigations to determine which

region in the chaperonins contains the amino acids involved in ATP-binding are currently in progress.

The possibility that the identified reversed dinucleotide binding site sequence is posttranslationally transpositioned in a reverse manner in the mature functional chaperonin seems unlikely but cannot be excluded, because all the chaperonin amino acid sequences are derived from cDNA sequences and not derived directly. Post-translational processing occurs in the major lectin seed storage protein from *Canavalia ensiformis* (jack bean), called concanavalinA (ConA), and involves the transposition of a 119 amino acid sequence from the carboxyterminus to the aminoterminal (Bowles *et al.*, 1986). This transportation is thought to be a transpeptidation event rather than a temporal separation and does not involve any major rearrangements in the tertiary folding of the single domain polypeptide chain. A mechanism for reversed transposition in proteins has not been reported, and if it does occur, must require a special means of breakage and ligation of the peptide backbone as simple peptide bond cleavage and religation would not be adequate. A more likely and interesting possibility is that a given binding site can be constructed from the same set of amino acids running in either direction along the polypeptide chain. This possibility is currently tested for the chaperonins and may turn out to have a more general occurrence. It is suggested that search programmes should be run in both directions in future to explore this interesting and unexpected possibility.

The challenge for the future is to determine what chaperonins recognize in other proteins to prevent the formation of incorrect, non-functional structures. One possibility is that chaperonins block certain folding pathways that lead to incorrect structure formation. Two enzymes have been identified that affect protein folding *in vitro* and are thought to be important *in vivo*. The cytoplasmic enzyme peptidyl-prolyl *cis-trans* isomerase (PPI) catalyzes the *cis trans* isomerisation of proline residues in a number of different proteins *in vitro* (Lang *et al.*, 1987) and the protein disulphide isomerase (PDI), present in the luminal space of the ER, catalyzes the rearrangement of disulphide bonds in polypeptides (Pain, 1987; Freedman, 1987). The PPI and PDI

activities appear to be important for the kinetics of protein folding but neither activity is involved in the specification of the final folded structure (see Ellis & Hemmingsen, 1989). Interestingly the PDI enzyme appears to be multifunctional, and catalyzes native disulphide formation, isomerization or reduction as a homodimer (Mr 114 000), but when assembled with 2 distinct  $\alpha$  subunits (Mr 64 000) it forms the tetrameric holoenzyme prolyl-4-hydroxylase, and catalyzes hydroxylation of proline residues in collagen polypeptides (Pain, 1987; Freedman, 1989). The reaction catalyzed by the chaperonins may well be similar to such isomerizations, and the association with other distinct polypeptides to form a 'new' protein complex with specificity for a particular cellular process is an interesting possibility. The observed requirement of additional factors such as a *groES*-like protein may be an indication of this possibility.

### 3.3.8 CONCLUDING REMARKS

The mechanism by which the chaperonins function is currently completely unknown and the existence of a class of proteins whose members appear to have similar but possibly not identical functions in all organisms may complicate the expression of such proteins in foreign hosts. The possible synthesis and assembly of the *Triticum aestivum* chloroplast  $\alpha$  chaperonin in *E.coli* is presented in the following chapter. Amino acid sequence comparison of members within the molecular chaperone family to identify regions of functional importance, together with the construction, by X-ray techniques of structures of domains, will provide further information to the function of these proteins. The development of an experimental system, in which either a polypeptide chain is folded or an oligomeric protein is assembled into a functionally active unit in the presence of a chaperonin *in vitro*, is urgently required to clarify the precise role of this group of molecular chaperones.

### 3.4 EXPRESSION OF THE *TRITICUM AESTIVUM* ALPHA CHAPERONIN AND THE RUBISCO LARGE AND SMALL SUBUNITS IN *ESCHERICHIA.COLI*

#### 3.4.1 INTRODUCTION

The failure to assemble an enzymically active *Triticum aestivum* Rubisco oligomer from its subunits expressed in *E.coli* suggests a requirement for plant-specific factor(s) involved in Rubisco assembly. The chloroplast chaperonin seems a good candidate for being such a plant-specific factor, since it is this protein to which newly-synthesized Rubisco large subunits bind in isolated chloroplasts (Barraclough & Ellis, 1980). The associated large subunits are thought to be released from the chloroplast chaperonin in an ATP-dependent fashion prior to their assembly into the holoenzyme (Milos & Roy, 1984; Milos *et al.*, 1985; Cannon *et al.*, 1986). The *E.coli* chaperonin mediates the assembly of cyanobacterial Rubisco subunits expressed in *E.coli* (Goloubinoff *et al.*, 1989) and is probably normally involved in a number of cellular processes that require protein-protein interactions (Section 3.3.7). Although the percentage amino acid sequence similarity between the bacterial and chloroplast chaperonins is 71% (Table 16), their predicted secondary structures are distinct (Section 3.3.4; Fig. 48) and the latter may reflect slightly different modes of action. Thus the failure of the wheat Rubisco subunits to assemble in *E.coli* may reflect the inability of the bacterial chaperonin to substitute for the chloroplast chaperonin. This suggests that the coexpression of the chloroplast  $\alpha$  chaperonin and the Rubisco large and small subunits in the same *E.coli* cell may be a way of rescuing Rubisco assembly; such experiments are described in this chapter. The *E.coli* chaperonin is of vital importance to the cell and the expression of a homologous protein such as the *Triticum aestivum* chloroplast  $\alpha$  chaperonin may lead to interference with the host protein in some of its essential functions. In this chapter studies on the synthesis, solubility and possible assembly of the wheat chloroplast  $\alpha$  chaperonin in *E.coli* are

described. Evidence is presented suggesting that the *E.coli* and chloroplast  $\alpha$  chaperonins form hybrid oligomeric structures (Section 3.4.2). Although wheat Rubisco subunits do not assemble into an enzymically active oligomer in *E.coli*, they are both soluble and stable (Bradley *et al.*, 1986; van der Vies *et al.*, 1986), so the possibility that these subunits are associated with the *E.coli* chaperonin was also investigated (Section 3.4.3).

### 3.4.2 EXPRESSION OF THE *TRITICUM AESTIVUM* CHLOROPLAST ALPHA CHAPERONIN IN *ESCHERICHIA COLI*

#### 3.4.2.1 Construction of expression vector pSV15

The *Triticum aestivum* cDNA sequence encodes the mature chloroplast  $\alpha$  chaperonin plus two amino acids of the presequence (Section 3.2.4; Fig. 20). To express the wheat chaperonin in *E.coli* a promoter, a ribosome binding site and a translation initiation codon (ATG) are required upstream of the cDNA sequence. As it is not known whether the expression of the chloroplast chaperonin will be detrimental to the *E.coli* cell, transcriptional control is essential and can be provided by the use of an inducible promoter. The ultimate aim is to coexpress the wheat chloroplast  $\alpha$  chaperonin and the wheat Rubisco large and small subunits in the same cell; it is therefore advantageous to be able to control the transcription of all three sequences in the same manner. The expression vectors containing the wheat Rubisco large and small subunit sequences under transcriptional control of the IPTG-inducible promoter of the *lac* gene, were already available (Fig 67; Bradley *et al.*, 1986). Thus the creation of a translational fusion between the  $\alpha$  chaperonin cDNA and a part of the *lacZ* gene, containing all the bacterial control sequences is a simple and rapid way to meet the requirements for expression of the  $\alpha$  chaperonin subunit in *E.coli*. The expressed  $\alpha$  chaperonin fusion protein will have a number of extra amino acids at the aminotermminus and it is assumed that this will not have an affect on the properties of

the chaperonin. Small aminoterminal extensions do not usually abolish the activity of the fusion proteins; for example, the addition of six amino acids to the amino terminus of the wheat Rubisco small subunit does not prevent its assembly with cyanobacterial Rubisco large subunits into an enzymically active hybrid oligomeric protein in *E.coli* (van der Vies, *et al.*, 1986).

The plasmid vector pUC9 contains the required bacterial control sequences and a polylinker region with suitable restriction enzyme sites (Vieira & Messing, 1982). The entire wheat chaperonin cDNA is present on the *EcoRI* fragment that was isolated from pSV8 (Fig. 10), followed by the end-filling of the recessed 3'ends. This blunt-end fragment was subsequently ligated into the *SmaI* site of pUC9 to generate pSV12. To fuse the wheat  $\alpha$  chaperonin sequence in the correct reading frame with the aminoterminal of  $\beta$ -galactosidase, a minimum of one extra nucleotide was needed between the translation initiation codon and the wheat cDNA sequence. This addition was achieved by digesting pSV12 with *BamHI*; the generated recessed 3'ends were then filled and the plasmid religated to give pSV15. Formation of the correct reading frame in pSV15 by the addition of four nucleotides was confirmed by sequencing (Fig. 60). Thus the resulting expression plasmid, pSV15, encodes the mature chloroplast  $\alpha$  chaperonin plus two amino acids of the presequence plus the first eighteen aminoterminal amino acids of  $\beta$ -galactosidase (Fig. 61).

#### 3.4.2.2 Synthesis of the *Triticum aestivum* chloroplast $\alpha$ chaperonin in *Escherichia.coli*

The expression of a foreign protein in an *E.coli* cell that is homologous to a host protein essential for cell viability may lead to several problems. The foreign protein may compete with the host protein in essential cellular processes, but if it is not able to perform the required function or only at a reduced efficiency, reduced cell growth or early cell death may result. On the other hand the *E.coli* cell may recognize the foreign homologue as an abnormal protein and degrade it rapidly. Rapid



Figure 60. Nucleotide sequence determination of the translational fusion region of pSV15

The nucleotide sequence of plasmid pSV12 (Panel B) and pSV15 (Panel A) was determined as described previously (Section 2.2.11) using the M13-20 primer. The different lanes represent the different nucleotides; A, adenine; C, cytosine; G, guanine and T, thymine. The correct translational fusion obtained by addition of four nucleotides (shown between the dotted lines) as a result of filling the recessed 3' ends of the generated *Bam*H1 site (see Section 3.4.2.1 and Figure 61). Amino acid residues are shown in the one-letter code as described by IUPAC-IUB (1970; 1984). The mature chloroplast  $\alpha$  chaperonin starts at D A K ...etc., whereas the preceding amino acids G and A form part of the presequence.

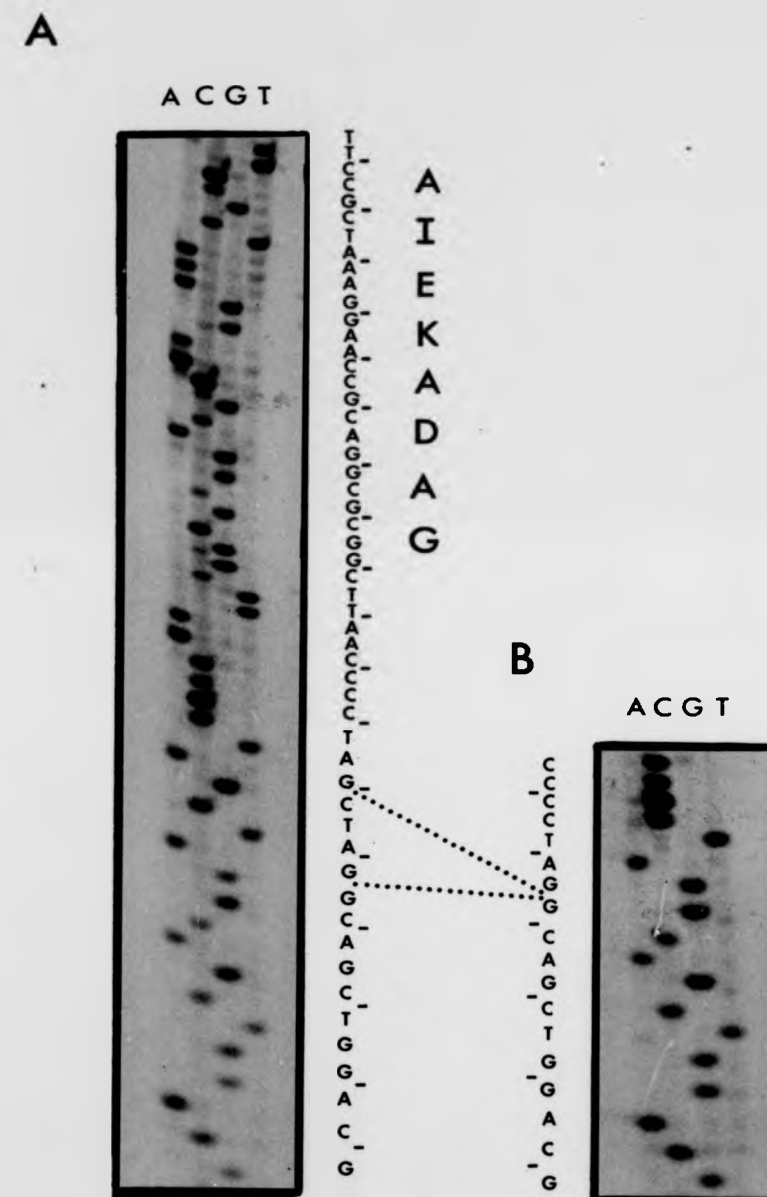
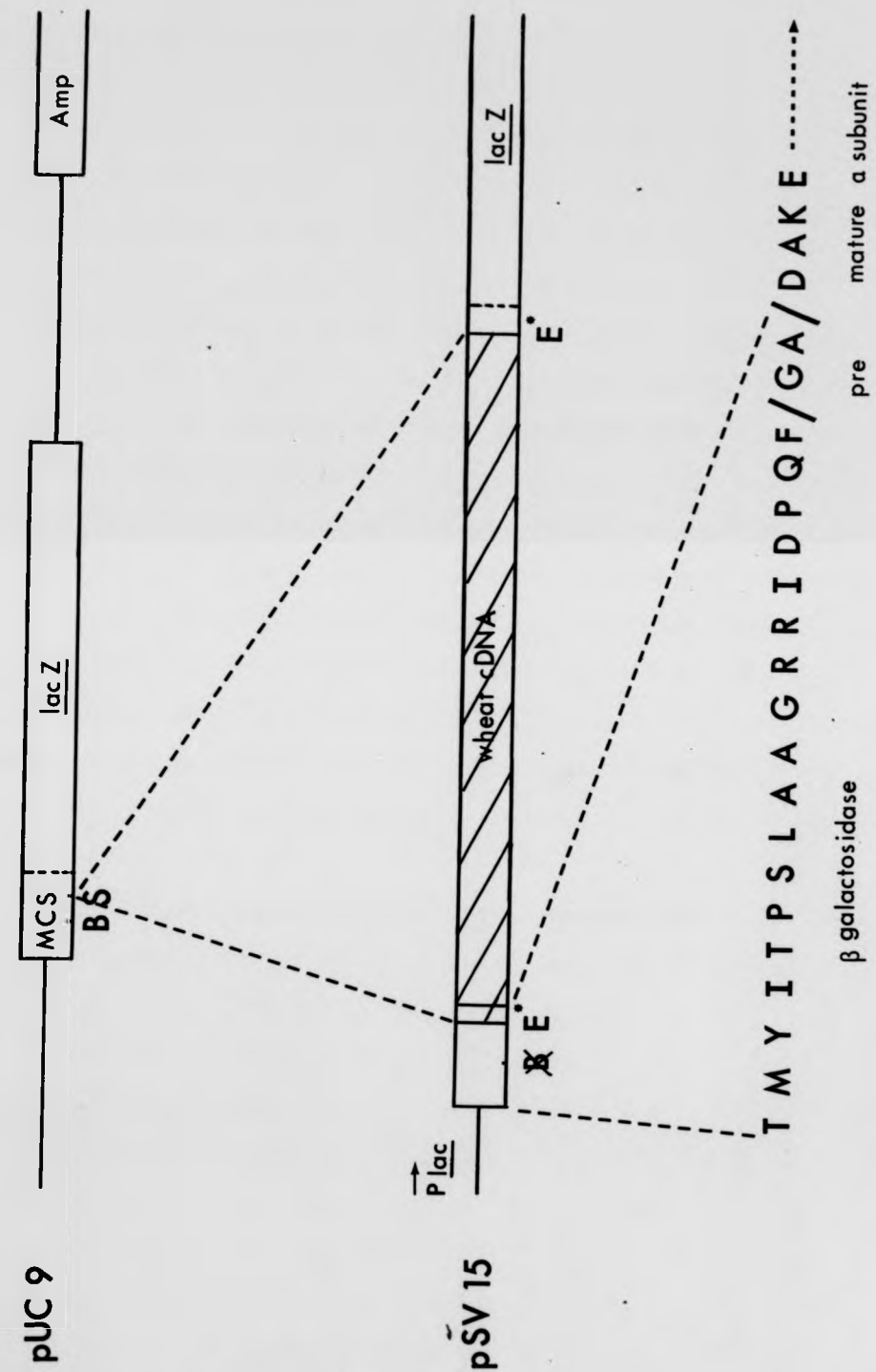


Figure 61. Structure of expression plasmid pSV15

Plasmid pSV8 (Figure 10) was digested with *Eco*RI followed by filling of the recessed 3' ends. The generated blunt end fragment containing the entire wheat cDNA sequence was separated on a 1% (w/v) agarose gel and subsequently isolated. This fragment was ligated into the plasmid vector pUC9 which had previously been digested with *Sma*I, and the ligation mixture was used to transform competent TG2 *E. coli* cells. The plasmid that contained the wheat cDNA fragment in an orientation which allows 5' → 3' transcription from the *lac* promoter was called pSV12 (not shown). The plasmid pSV12 was digested with *Bam*HI, the generated 3' ends were filled and the religated DNA used to transform competent *E. coli* TG2 cells. The new construct, named pSV15, contains a translational fusion of the chloroplast  $\alpha$  chaperonin sequence and the aminotermminus of  $\beta$ -galactosidase. The junction of  $\beta$ -galactosidase and the  $\alpha$  chaperonin is marked by a slash (/). The amino acid residues are written in the one-letter code (IUPAC-IUB, 1970; 1984). The *Eco*RI restriction enzyme that has been end-filled (E\*) and the *Bam*HI site (B) destroyed are indicated. The arrow indicates the direction of transcription.



degradation of the foreign homologue would also reduce its possible interference with host cellular processes and thus allow the cell to survive.

The synthesis of aberrant and fragmented proteins in *E.coli* increases the activity of the bacterial proteolytic system involved in the degradation of such proteins (Goff & Goldberg, 1985). The degradation of abnormal proteins by the *E.coli* cell is not a trivial problem as is demonstrated by the unrelated bacteriophages T4, T5 and T7, which all contain the *pin* gene, whose product inhibits the degradation of abnormal proteins, but which does not affect the turnover of normal *E.coli* proteins (Simon *et al.*, 1978; Simon *et al.*, 1983). The early synthesis of the *pin* gene product is thought to be an evolutionary adaptation which protects phage proteins from being recognized and degraded as abnormal proteins by the proteases of the host cell.

On the other hand the wheat  $\alpha$  chaperonin may be able to complement the normal functions of the *E.coli* chaperonin in which case little or no effect on the host cell is anticipated. The  $\alpha$  chaperonin is one of the two types of polypeptide that make up the chloroplast chaperonin oligomer (Hemmingsen & Ellis, 1986) and shows about 50% amino acid identity with the  $\beta$  chaperonin (Table 15). The percentage amino acid identity between the wheat  $\alpha$  chaperonin and the *E.coli* chaperonin is 46% (Table 15) so it may not be unexpected if the latter two chaperonins are able to form hybrid structures.

To examine the expression of the wheat chloroplast  $\alpha$  chaperonin, *E.coli* TG2 cells containing the plasmid pSV15 were grown in the absence of IPTG to mid-log phase. At this stage IPTG was added to a final concentration of 0.5 mM and the growth of the cells was measured. No change in growth rate was observed compared to the wild type TG2 cells and a doubling time of 20 min was determined (data not shown). The cells were harvested at stationary phase and soluble protein extracts were separated by SDS polyacrylamide gel electrophoresis, followed by immunoblotting. To detect the wheat chloroplast  $\alpha$  chaperonin a pea chloroplast chaperonin antibody preparation was used, from which all the *E.coli*-reacting antibodies had been removed (Section 2.2.13.2). The *E.coli* chaperonin was detected using specific antibodies

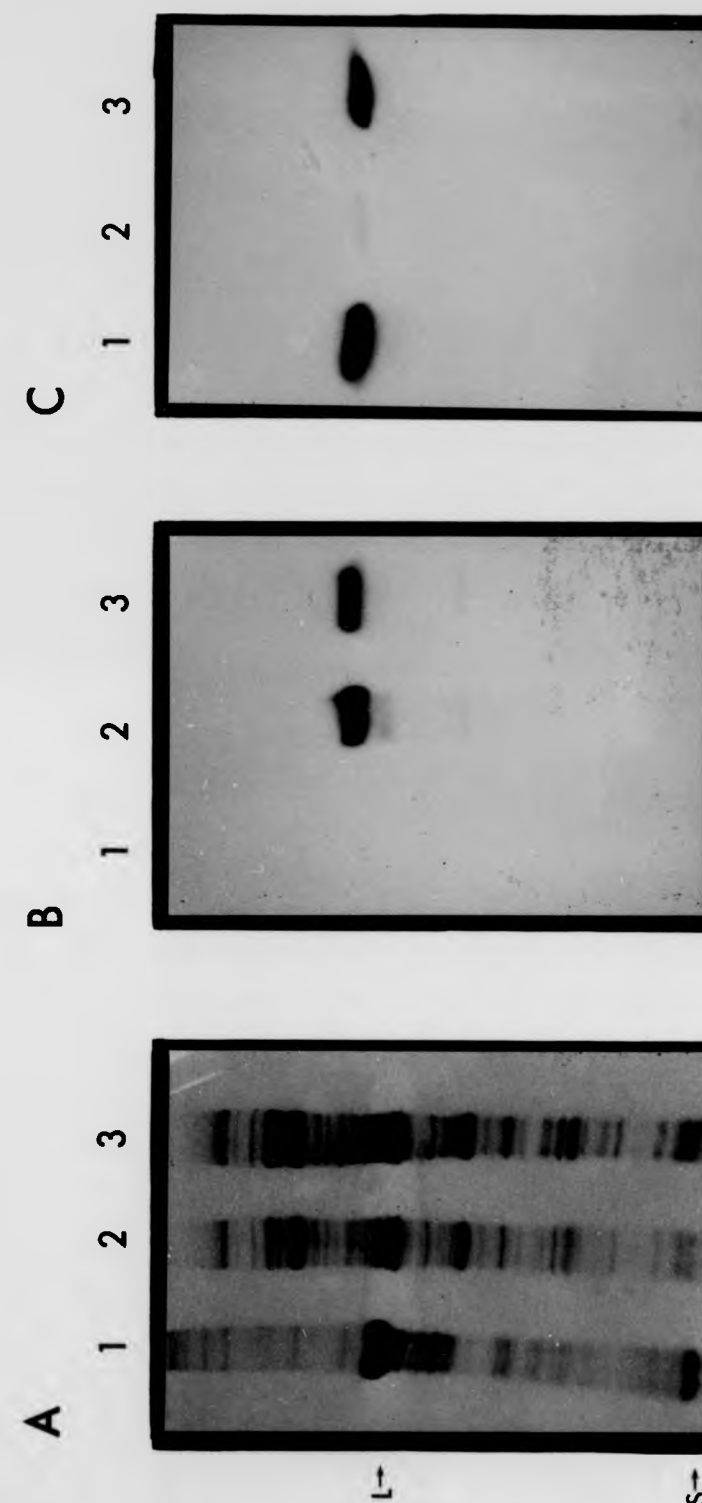
raised against this protein. The chloroplast  $\alpha$  chaperonin is detectable in extracts of *E.coli* cells containing plasmid pSV15 (Fig. 62C, lane 3). The control (Fig. 62C, lane 2) shows that virtually all the *E.coli* chaperonin antibodies had been removed from the chloroplast chaperonin antibody preparation. The synthesis and solubility of the *E.coli* chaperonin is not affected by the expression of the chloroplast  $\alpha$  chaperonin (Fig. 62B, cf lane 2 and lane 3). It is concluded that the wheat chloroplast  $\alpha$  chaperonin is synthesized as a soluble stable protein in *E.coli* and that this synthesis affects neither the synthesis of the *E.coli* chaperonin nor cell growth.

#### 3.4.2.3 Assembly of the *Triticum aestivum* chloroplast $\alpha$ chaperonin into homologous and heterologous oligomers in *Escherichia coli*

The purified chaperonin oligomer from *Pisum sativum* and *Hordeum vulgare* chloroplasts consists of two types of polypeptide,  $\alpha$  and  $\beta$ , which are present in equal amounts in the purified protein (Hemmingsen & Ellis, 1986; Musgrove & Ellis, 1986; Musgrove *et al.*, 1987; Johnson, 1987), while the *E.coli* chaperonin is purified as a homo oligomer (Hendrix, 1979). There is evidence that both oligomers are 14-mers (Hendrix, 1979; Pushkin *et al.*, 1982). To examine the physical state of the *E.coli* and wheat chloroplast chaperonins in *E.coli*, soluble protein extracts of *E.coli* TG2 cells and TG2 cells containing pSV15 were separated on a polyacrylamide gel under reducing non-denaturing conditions followed by immunoblotting with either *E.coli* chaperonin-specific antibodies or wheat  $\alpha$  chaperonin-specific antibodies. The mobility of proteins on non-denaturing gels depends on the ratio of charge to mass (Hedrick & Smith, 1968). The *E.coli* chaperonin in wild type *E.coli* TG2 cells was detected in three different positions on the gel (Fig. 63B, lane 2). The chaperonin detected near the top of the polyacrylamide gel (Fig. 63B, lane 2) is also seen on the Coomassie stained gel after immunoblotting as an identically shaped band (cf lane 2, Fig. 63A and Fig. 63B). Two faster mobility forms of the *E.coli* chaperonin are also detected

Figure 62. Synthesis of the *Triticum aestivum* chloroplast  $\alpha$  chaperonin in *E.coli*

To examine chloroplast chaperonin synthesis, *E.coli* TG2 cells containing the plasmid pSV15 were grown at 37°C in 400 ml of YT medium (Section 2.2.2) with 100  $\mu$ g/ml ampicillin and 0.5 mM IPTG to induce transcription from the *lac* promoter. In addition TG2 cells were grown under identical conditions in identical medium without ampicillin and IPTG. At  $A_{600} = 1.2-1.4$ , the cells were harvested by centrifugation in a MSE 21 centrifuge at 6000 g for 15 min at 4°C, washed in 100 mM Tris-HCl pH 8.0 containing 20 mM MgCl<sub>2</sub>, 1mM EDTA, 1 mM 2-mercaptoethanol and 1 mM PMSF (TEM), weighed and resuspended in an equal amount of TEM buffer. A soluble protein extract was prepared as described in Section 2.2.20.1. Wheat leaves were used to prepare an extract as described in Section 2.2.20.2. Extracts were analyzed on a 15% (w/v) polyacrylamide gel containing SDS, followed by immunoblotting. The gel was stained with Coomassie brilliant blue (Panel A). The nitrocellulose filters were either incubated with a purified preparation of chloroplast chaperonin antibodies raised against the purified oligomer from *Pisum sativum* (Panel C) or with antibodies raised against the purified *E.coli* chaperonin (Panel B). Antibodies were detected using the radiolabelled detection method (Section 2.2.14.2.2) Lane 1, wheat leaf extract; lane 2, extract of *E.coli* TG2 cells; lane 3, extract of *E.coli* TG2 cells containing pSV15. The positions of the Rubisco large (L) and small (S) subunits on the polyacrylamide gel are indicated.

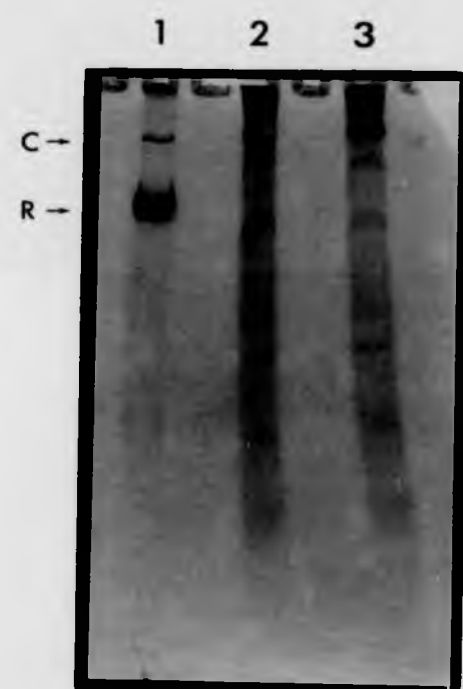


**Figure 63. Analysis of the *Triticum aestivum* chloroplast  $\alpha$  chaperonin in *E.coli* extracts under non-denaturing conditions**

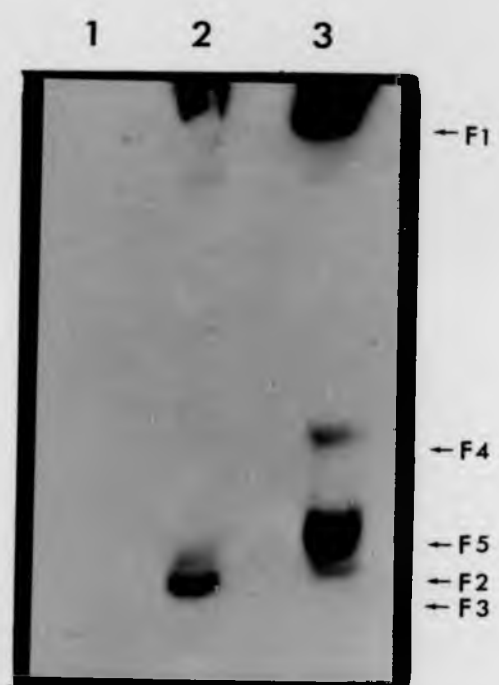
Soluble protein extracts of *E.coli* TG2 cells and TG2 cells containing pSV15 were prepared from cells grown as described in the legend to Figure 62 with the difference that the extraction buffer used was composed of 100 mM Tris-HCl pH 8.0 containing 20 mM MgCl<sub>2</sub>, 10 mM Na<sub>2</sub>CO<sub>3</sub>, 1 mM EDTA, 1 mM 2-mercaptoethanol and 1 mM PMSF. The wheat leaf extract was prepared using the same buffer. The protein extracts were analyzed on a 4% to 15% polyacrylamide linear gradient gel under non-denaturing and reducing conditions (Section 2.2.12.2.1), followed by immunoblotting and incubation with either specific antibodies against the chloroplast chaperonin or specific antibodies raised against the *E.coli* chaperonin (see also the legend to Figure 62). The polyacrylamide gels were stained with Coomassie brilliant blue after immunoblotting. Panel A1, stained gel of Panel B; Panel B, autoradiograph after incubation with the *E.coli* chaperonin antibodies; Panel A2, stained gel of Panel C; Panel C, autoradiograph after incubation with chloroplast chaperonin antibodies. Lane 1, wheat leaf extract; lane 2, extract of TG2 cells; lane 3; extract of TG2 cells containing pSV15. The wheat chaperonin (C) and the Rubisco holoenzyme (R) are indicated. F1 to F5 refer to the different forms of the *E.coli* and the chloroplast chaperonin polypeptides



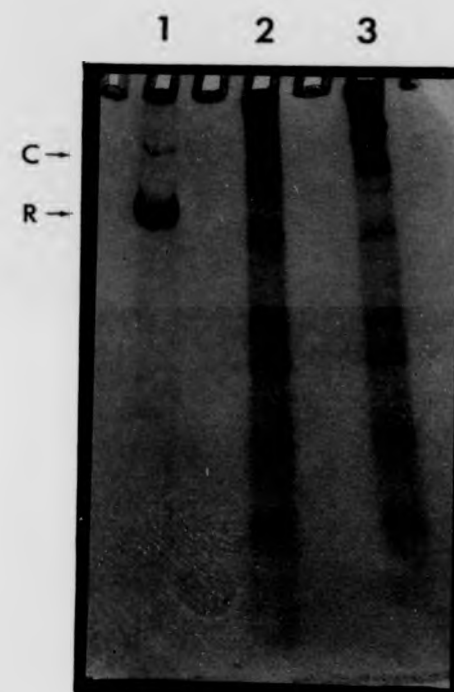
A1



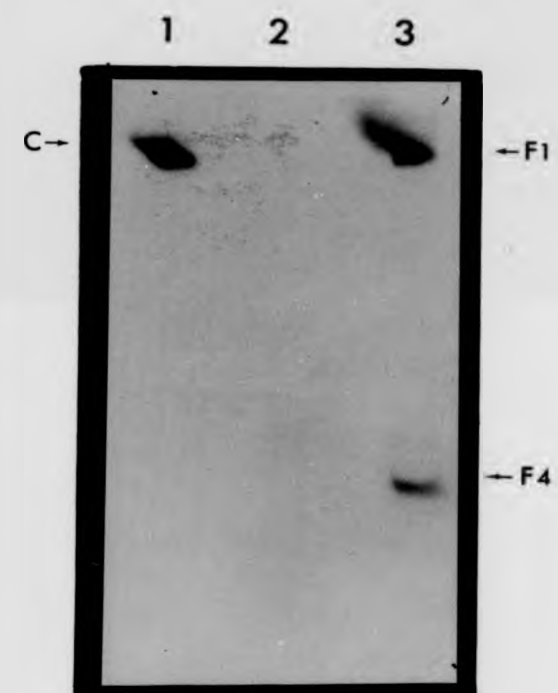
B



A2



C





(Fig. 63B, lane 2, bands F2 and F3), but whether they represent monomers is not known. Multiple forms of the chaperonin in extracts of *E.coli* have not been reported before, and may represent functionally different forms of the protein. The amount of chaperonin in the *E.coli* cells is about 1% of the total amount of protein under normal physiological conditions (Neidhardt *et al.*, 1981). The concentration of the chaperonin in the *E.coli* extracts analyzed here is about 0.2-0.3 mg/ml, probably 50 to 100-fold lower than the concentration *in vivo*. If the *E.coli* chaperonin oligomer dissociates reversibly, as has been demonstrated for the chloroplast chaperonin, it is expected that an increased protein concentration will favour association (See also Section 3.3.7.7). The expression of the wheat chloroplast  $\alpha$  chaperonin in *E.coli* results in the disappearance of the most mobile form of the *E.coli* chaperonin (Fig. 63B, lane 3, band F3) and the appearance of two additional bands that cross-react with the *E.coli* chaperonin antibodies (Fig. 63B, lane 3, band F4 and F5). Whether these additional forms consist of heterologous *E.coli* and wheat  $\alpha$  chaperonins remains to be determined.

When identical *E.coli* extracts of cells containing pSV15 are treated with antibodies against the chloroplast chaperonin, two cross-reacting proteins are observed (Fig. 63C, lane 3). The lower mobility form runs at a position almost similar to the oligomeric chaperonin in a wheat leaf extract (Fig. 63A2, cf lane 1 and 3). This form either represents the wheat  $\alpha$  chaperonin homo-oligomer or is a hybrid protein of *E.coli* and wheat  $\alpha$  chaperonin polypeptides.

The higher mobility form which cross-reacts with the wheat chaperonin antibodies (Fig. 63C, lane 3, band F4) runs at a position similar to a form that is recognized by the antibodies against the *E.coli* chaperonin (Fig. 63B, lane 3, band F4). This form (F4) is observed only in extracts of *E.coli* that contain both the *E.coli* and the wheat chloroplast  $\alpha$  chaperonins; this correlation suggests strongly that these two types of chaperonin assemble into a heterologous oligomeric protein complex. The possibility that other identified forms contain both *E.coli* and wheat chaperonin polypeptides but are assembled in such a way that one type is masked by the other type

and therefore the oligomer is recognized by one specific antibody only, is unlikely but can not be excluded. Whether the observed hybrid oligomers are also present *in vivo* and are functionally active remains to be seen.

#### 3.4.2.4 Analysis of the *Escherichia coli* and *Triticum aestivum* chloroplast $\alpha$ chaperonins in *Escherichia coli* extracts by velocity sedimentation centrifugation

To determine the composition of the observed chaperonin protein structures, a soluble protein extract of *E.coli* cells harbouring pSV15 was separated by sucrose density gradient centrifugation, followed by analysis of the fractions on SDS polyacrylamide gels and immunoblotting, using either an *E.coli* or a wheat chloroplast chaperonin-specific antibody preparation. It can be seen from Fig. 64A that the *E.coli* chaperonin polypeptides are present in a number of the sucrose gradient fractions but are not assembled into oligomeric complexes of the size expected for a 14-mer ( $M_r$  about 800 000). The wheat chloroplast  $\alpha$  chaperonin polypeptides are present in exactly the same fractions as the *E.coli* chaperonin polypeptides (Fig. 64B). This identical distribution suggests that either the two chaperonins assemble into homooligomeric structure that behave identically on a sucrose gradient or, more likely, that heterologous protein oligomers are formed with  $M_r$ s varying from 100 000 to 400000.

Earlier studies by Murialdo (1979), who analyzed protein extracts of *E.coli* mutants infected with mutant  $\lambda$  bacteriophages, showed that the *E.coli* chaperonin in crude extracts sediments at a position expected for the 14 subunit oligomer, while the partially purified *E.coli* chaperonin also sediments as a 14-mer (Hendrix, 1979). However the extraction buffer used in these studies was different from the one used here as the latter did not contain any  $Mg^{2+}$  ions. To examine the possible effect of magnesium ions on the sedimentation of the protein complexes, *E.coli* cells containing pSV15 were lysed in 100 mM Tris-HCl pH 8.0 containing 10 mM  $MgCl_2$ , 5 mM EDTA, 5 mM 2-mercaptoethanol and 1 mM PMSF (i.e. identical to the extraction

Figure 64. Analysis of the *Triticum aestivum* chloroplast  $\alpha$  chaperonin polypeptides in *E.coli* extracts by velocity sedimentation centrifugation in the absence of  $Mg^{2+}$  ions

A soluble protein extract in 100 mM Tris-HCl pH 8.0, containing 5 mM EDTA, 1 mM 2-mercaptoethanol and 1 mM PMSF, of *E.coli* TG2 cells containing pSV15 was analyzed on a linear sucrose density gradient (7-25%). The fractions were analyzed as described in the legend to Figure 62. Panel A shows the autoradiograph after incubation with the *E.coli* chaperonin specific antibodies and Panel B shows the autoradiograph after incubation with the chloroplast chaperonin antibodies. The extreme right lanes in (A) and in (B) contain a *E.coli* TG2 cell extract (C1) and a wheat leaf cell extract (C2) respectively, which was used as markers for the *E.coli* and wheat chaperonins. Purified  $\beta$ -galactosidase (Mr 465 000) was used as as internal marker; aldolase (Mr 120 000) and ovalbumine (Mr 43 000) were used to calibrate the gradient. Numbers at the top indicate relative molecular masses in terms of kDa and the positions of the *E.coli* (EC) and the chloroplast chaperonins (CC) are indicated.

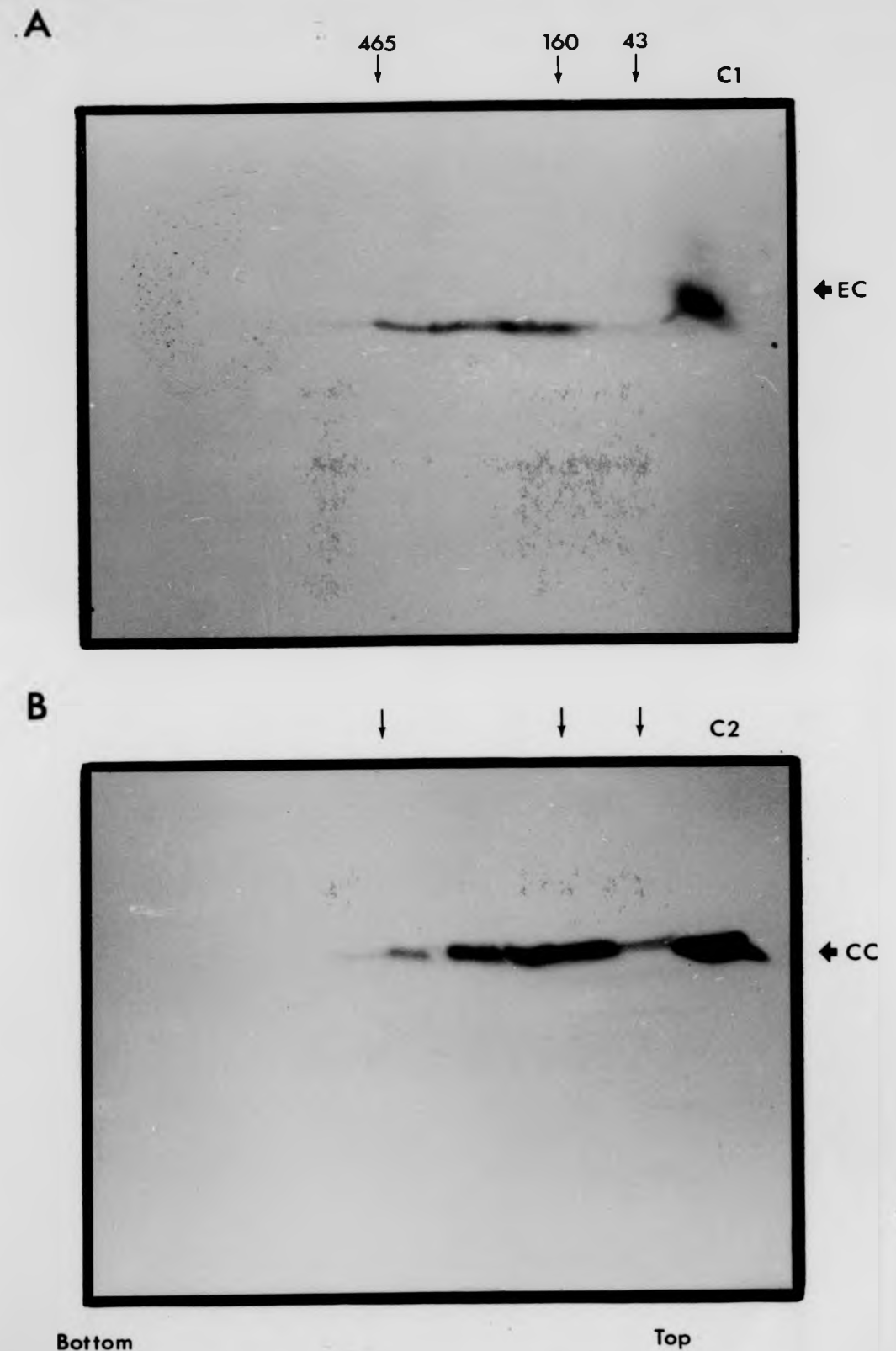
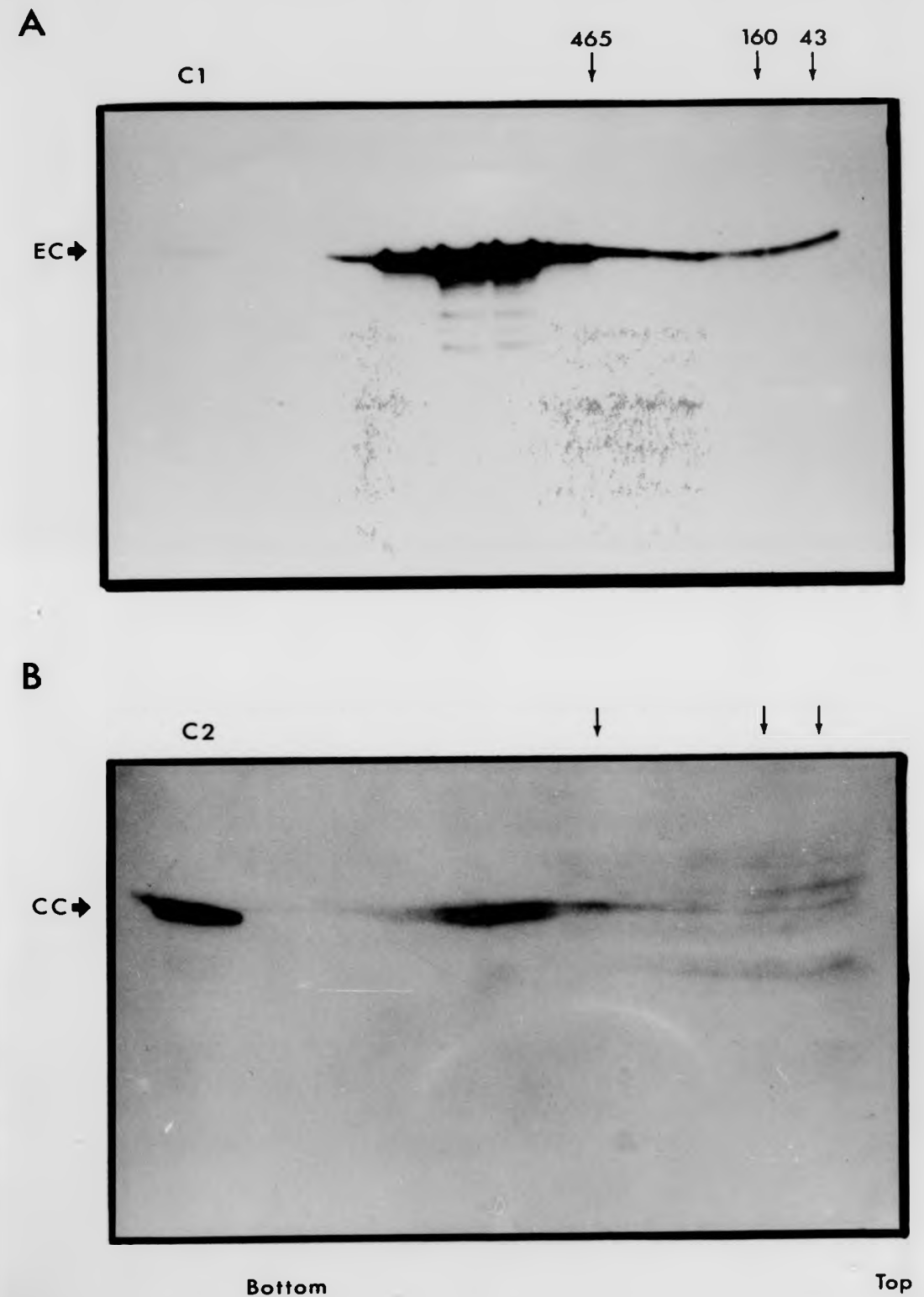


Figure 65. Analysis of *Triticum aestivum* chloroplast  $\alpha$  chaperonin polypeptides in *E.coli* extracts <sup>by</sup> velocity sedimentaion centrifugation in the presence of  $Mg^{2+}$  ions

A soluble protein extract in 100 mM Tris-HCl pH 8.0 containing 10 mM  $MgCl_2$ , 5 mM EDTA, 1 mM 2-mercaptoethanol and 1 mM PMSF, of *E.coli* TG2 cells containing pSV15 was fractionated on sucrose gradients and analyzed as described in the legend to Figure 64. The autoradiograph after incubation with *E.coli* chaperonin-specific antibodies (Panel A) and chloroplast chaperonin-specific antibodies (Panel B) are shown. Markers; *E.coli* TG2 and wheat leaf extracts, lane C1 and C2 respectively. Calibration proteins are identical to the ones described in Figure 64.



buffer used before but with  $\text{MgCl}_2$  added). The soluble protein extract was separated on a sucrose density gradient and the fractions analyzed by SDS polyacrylamide gel electrophoresis and immunoblotting using the previously described specific antibodies. The result shows that the majority of the *E. coli* chaperonin protein migrates at a position indicating a  $M_r > 465\,000$ , which probably represent the assembled 14-subunit oligomer (Fig. 65A). A small amount of the *E. coli* chaperonin migrates nearer the top of the gradient indicating monomers and/or dimers. Some proteolytic degradation is observed since minor lower molecular weight polypeptides are present in the gradient peak fractions (Fig. 65A). The presence of at least two forms of the *E. coli* chaperonin, in the presence of  $\text{Mg}^{2+}$  ions is comparable to results obtained using polyacrylamide gel electrophoresis (Fig. 63B, lane 2). It is possible that form F1 in Fig. 62B is oligomeric, while forms F2 and F3 have lower molecular weights.

Interestingly the effect of  $\text{Mg}^{2+}$  ions on the distribution of the wheat chloroplast  $\alpha$  chaperonin across the sucrose gradient is identical to that observed for the *E. coli* chaperonin (cf. Fig. 65A and 65B). The identical new distribution of both the chaperonins in the presence of  $\text{Mg}^{2+}$  ions supports the idea that heterologous protein complexes are present. The presence of  $\text{Mg}^{2+}$  ions may create a favourable environment for the chaperonin polypeptides to assemble into large oligomeric protein complexes by affecting the equilibrium between the different forms. This requirement of  $\text{Mg}^{2+}$  ions for the oligomeric chaperonin assembly in *E. coli* extracts needs further investigation since the purified *E. coli* chaperonin dissociates in the presence of MgATP (A.A. Gatenby, personal communication); while the reversible dissociation of the chloroplast chaperonin in stromal extracts of *P. sativum* is dependent on the concentration MgATP but not on the concentration of  $\text{Mg}^{2+}$  ions alone (Hemmingsen & Ellis, 1986; R.J. Ellis, personal communication).



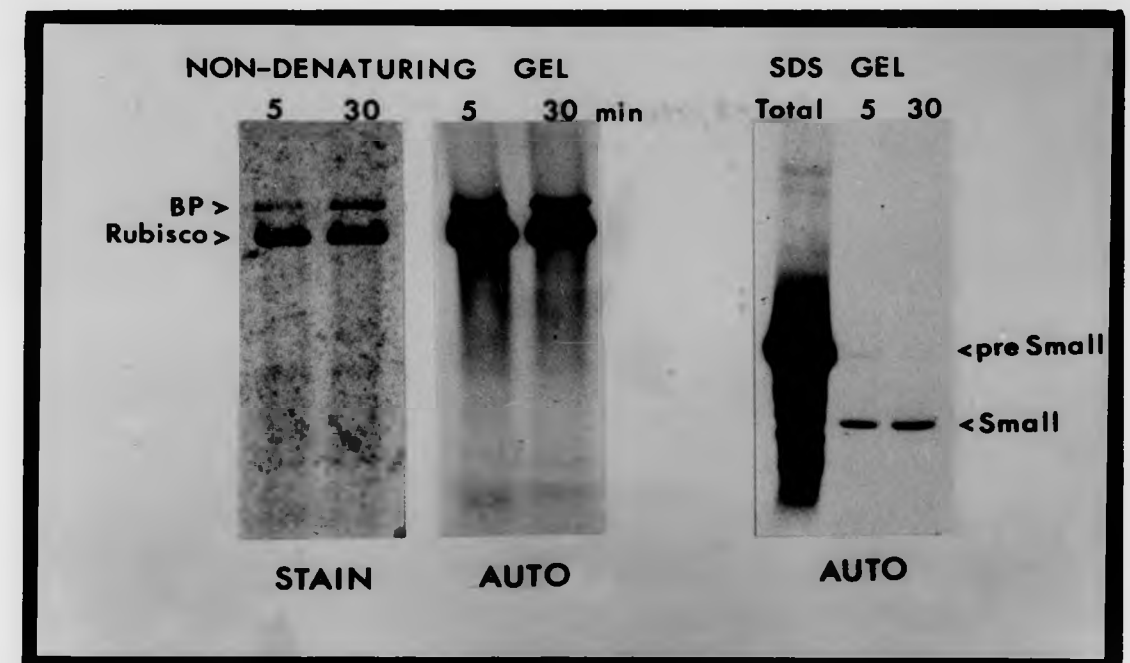
### 3.4.3 A COMPARISON OF THE FORM OF THE *TRITICUM AESTIVUM* RUBISCO SUBUNITS EXPRESSED IN *ESCHERICHIA COLI* WITH *PISUM SATIVUM* RUBISCO SUBUNITS IN ISOLATED PROTOPLASTS

The assembly pathway of the hexadecameric Rubisco enzyme in the chloroplasts of plants is not completely understood. However a better understanding of this process seems essential if the ultimate aim is to assemble an active plant Rubisco enzyme in an *E.coli* cell so that genetic engineering can commence. The synthesis of the wheat Rubisco large and small subunits in *E.coli* is clearly not sufficient to assemble an active Rubisco oligomer (Bradley *et al.*, 1986). In isolated chloroplasts of *Pisum sativum* newly-synthesized Rubisco large subunits are associated with the chloroplast chaperonin (Barracough & Ellis, 1980). The function of the chaperonin is suggested to be, to hold the large subunits in a conformation suitable for interaction with the small subunits and this association may be an essential step in the assembly of the Rubisco holoenzyme. So it might be expected that a transient association of small subunits with the chaperonin as well as large subunits occurs as part of the assembly process. Such an association was detected when labelled small subunit precursor was supplied to intact isolated *P.sativum* chloroplasts, followed by analysis of the stromal fraction on a non-denaturing polyacrylamide gel. As can be seen in Fig. 66 the majority of the imported small subunits assembles rapidly into the Rubisco holoenzyme, but a small proportion migrates with the chaperonin oligomer. Whether the small subunits are attached to the chaperonin directly or to the bound large subunits is not known. These observations suggest that the association of both large and small subunits with the chaperonin may be an essential step in the assembly of the Rubisco holoenzyme in plant chloroplasts. Identical observations have been made independently by Gatenby *et al.* (1988).

During the course of this work it became known that the *E.coli* chaperonin is required for the assembly of the cyanobacterial Rubisco subunits synthesized in *E.coli* (Goloubinoff *et al.*, 1989). This observation raises the question whether the wheat

Figure 66. Association of imported Rubisco small subunits with the chloroplast chaperonin

Labelled small subunit precursor from *Pisum sativum* was made in a wheat-germ extract containing  $^{35}\text{S}$ -methionine from mRNA transcribed from an SP6 plasmid containing the pea small subunit precursor cDNA (kindly provided by N-H. Chua). The labelled precursor was incubated with isolated intact *P. sativum* chloroplasts in the presence of excess unlabelled methionine and light as an energy source. Samples were removed at 5 and 30 min and the stromal extracts analyzed by electrophoresis on a non-denaturing polyacrylamide gradient gel [4-30 % (w/v)] and a 15 % (w/v) polyacrylamide gel containing SDS. The gels were stained with Coomassie brilliant blue and subjected to autoradiography. Symbols: Total, complete translation products from SP6 plasmid; preSmall, precursor to the Rubisco small subunit; Small, Rubisco mature small subunit; BP, chaperonin; Rubisco, Rubisco enzyme.





Rubisco subunits that fail to assemble into an enzymically active enzyme (Bradley *et al.*, 1986) are associated with the *E.coli* chaperonin. It was decided to examine this possibility. Expression vectors to synthesize the wheat Rubisco large and small subunit were already available. The chloroplast large subunit gene is present in the M13 phage vector M68 under transcriptional control of the  $\beta$ -galactosidase gene promoter (Fig. 67A; Bradley *et al.*, 1986). The wheat Rubisco mature small subunit sequence is present in the expression plasmid vector pDB85 (kindly constructed by D. Bradley), as a translational fusion with a small part of the 5' end sequence of the *lacZ* gene (Fig. 67B), and this expression plasmid is similar to the earlier described M72-phage expression vector (van der Vies *et al.*, 1986). The synthesized small subunit fusion protein has been shown to be functionally active when assembled with cyanobacterial large subunits in *E.coli* (van der Vies *et al.*, 1986).

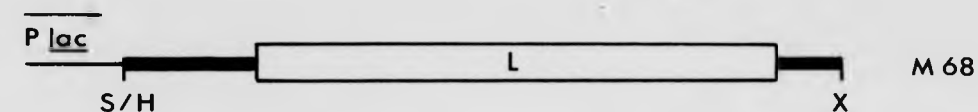
To examine the form of the wheat Rubisco subunits synthesized in *E.coli*, soluble protein extracts of either TG2 cells infected with M68, or TG2 cells containing pDB85 and infected with M68, were analyzed by non-denaturing polyacrylamide gel electrophoresis and immunoblotting. Specific antibodies against the *E.coli* chaperonin and the wheat Rubisco enzyme were used to detect the different polypeptides. The *E.coli* chaperonin is present as an oligomeric protein of probably 14 subunits in all four of the *E.coli* extracts as expected; however a higher mobility form is also observed in all four extracts (Fig. 68A), confirming earlier observations (Section 3.4.2.3, Fig. 63B). As expected, no cross-reactivity with any wheat leaf proteins is observed using the *E.coli* chaperonin antibodies (Fig. 68A., lane 1). The wheat Rubisco large subunits were detected at a position similar to the *E.coli* chaperonin oligomer suggesting that the two polypeptides are associated (Fig. 68B, lane 3). No cross-reacting polypeptides are observed in the *E.coli* extract of cells containing pDB85, suggesting that either the wheat Rubisco small subunits are not associated with the *E.coli* chaperonin, or are unstable or not synthesized. The latter two explanations were shown to be incorrect however, since stable soluble small subunits were detected when an aliquot of the same *E.coli* extract was analyzed under

Figure 67. Structure of the *Triticum aestivum* Rubisco subunit expression vectors

Panel A shows the structure of the wheat Rubisco large subunit expression vector M68 and was previously described by Bradley *et al.* (1986). A 1.9 kb *HincII-XbaI* DNA fragment of the entire wheat large subunit gene (T.A Dyer and C.M. Bowman, personal communication; Terachi *et al.*, 1987) was isolated from pTac 39 (Howe *et al.*, 1982) and ligated into M13mp18-RF (Norrander *et al.*, 1983; Yanish-Perron *et al.* 1985) which had been digested with *SmaI* and *XbaI*. Transcription of the large subunit gene occurs from the *lac* promoter (*Plac*) and the direction of transcription is indicated by the arrow. Symbols; S, *SmaI*; H, *HincII*; X, *XbaI* and L, Rubisco large subunit.

Panel B shows the structure of the wheat Rubisco small subunit expression vector pDB85 which was kindly constructed by D. Bradley and is similar to the previously described M72 vector (van der Vies *et al.* 1986). M72 was made by isolation of a 600 bp *BalI - PstI* fragment from W9 (Broglie *et al.*, 1983 and kindly provided by N-H. Chua), containing the DNA sequence coding for the mature small subunit lacking the first four amino acids at the aminotermminus, and ligation of this fragment into M13mp18-RF (Norrander *et al.*, 1983; Yanish-Perron *et al.*, 1985) that had been cleaved with *SmaI* and *PstI*. To construct pDB85, the *EcoRI - PstI* fragment isolated from M72-RF, containing the  $\beta$ -galactosidase/mature small subunit translational fusion, was ligated into pUC18 that had been cleaved with *EcoRI* and *PstI*. Symbols: P, *PstI*; B, *BalI*; S, *SmaI*; PS, presequence; mature S, Rubisco mature small subunit. Transcription occurs from the *lac* promoter (*Plac*) in the direction indicated by the arrow, whereas translation initiated at the *lacZ* AUG with the reading frame being maintained into the small subunit sequence. The aminoterminal sequences of the small subunit fusionprotein (pDB85) and the wild-type (WT) small subunit are presented in the one-letter code. The fusion junction is marked with a slash (/).

A



B

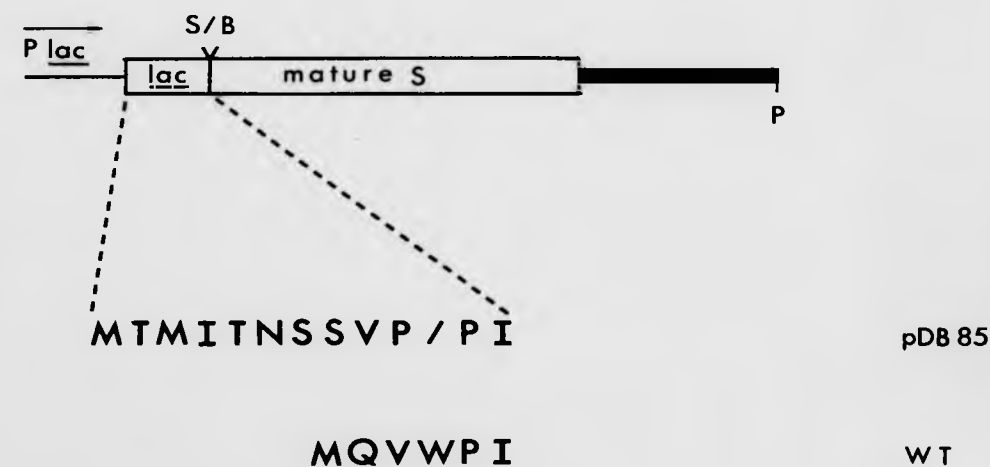
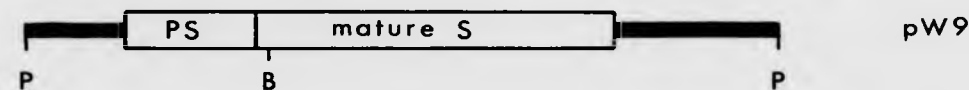
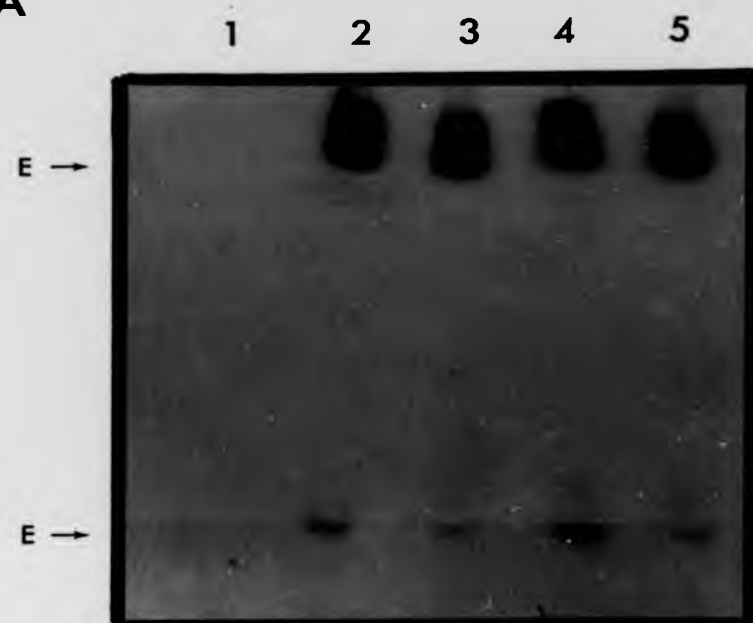


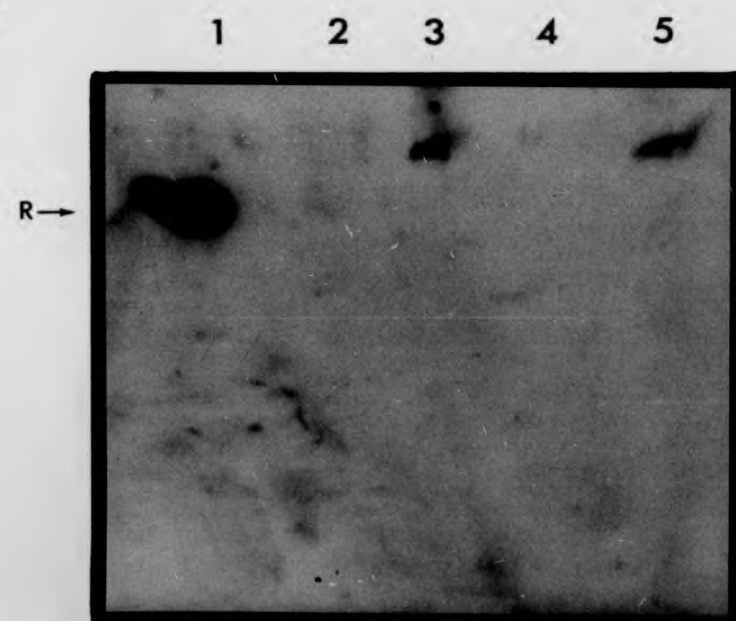
Figure 68. Association of *Triticum aestivum* Rubisco large subunits with the *E.coli* chaperonin in *E.coli* extracts

Soluble protein extracts in 100 mM Tris-HCl pH 8.0, containing 20 mM MgCl<sub>2</sub>, 10 mM Na<sub>2</sub>CO<sub>3</sub>, 1 mM EDTA, 1 mM 2-mercaptoethanol and 1 mM PMSF, of *E.coli* TG2 cells (lane 2), *E.coli* TG2 cells infected with M68 (lane 3), *E.coli* TG2 cells containing pBD85 (lane 4) and *E.coli* TG2 cells containing pDB85 and infected with M68 (lane 5) were separated by electrophoresis on non-denaturing polyacrylamide gradient gels [4-15% (w/v)] followed by immunoblotting and incubation with specific *E.coli* chaperonin antibodies (Panel A) and specific Rubisco antibodies (Panel B). Lane 1 contains a wheat leaf extract used as a marker for the Rubisco holoenzyme. Symbols: E, *E.coli* chaperonin; R, Rubisco. Panel C shows the autoradiograph of the protein extracts analyzed on a 15% polyacrylamide gel containing SDS, followed by immunoblotting and incubation with wheat Rubisco specific antibodies. The Rubisco large (L) and small (S) subunits are indicated.

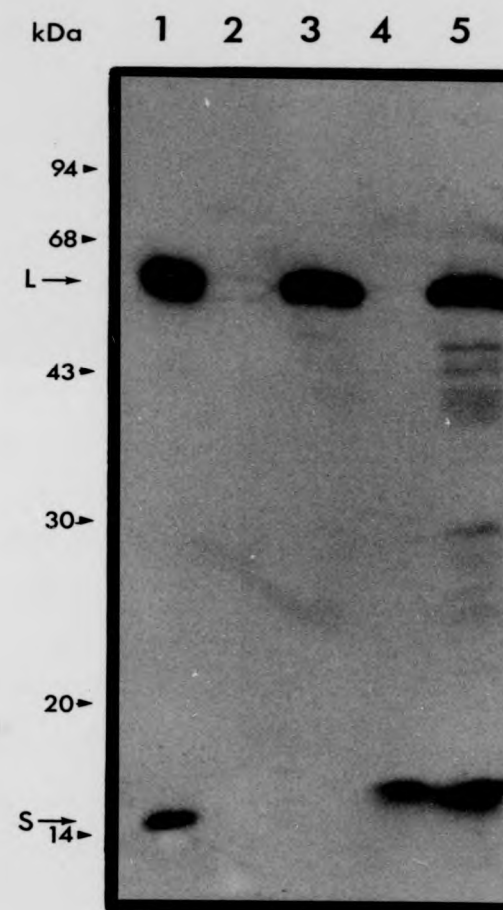
**A**



**B**



**C**



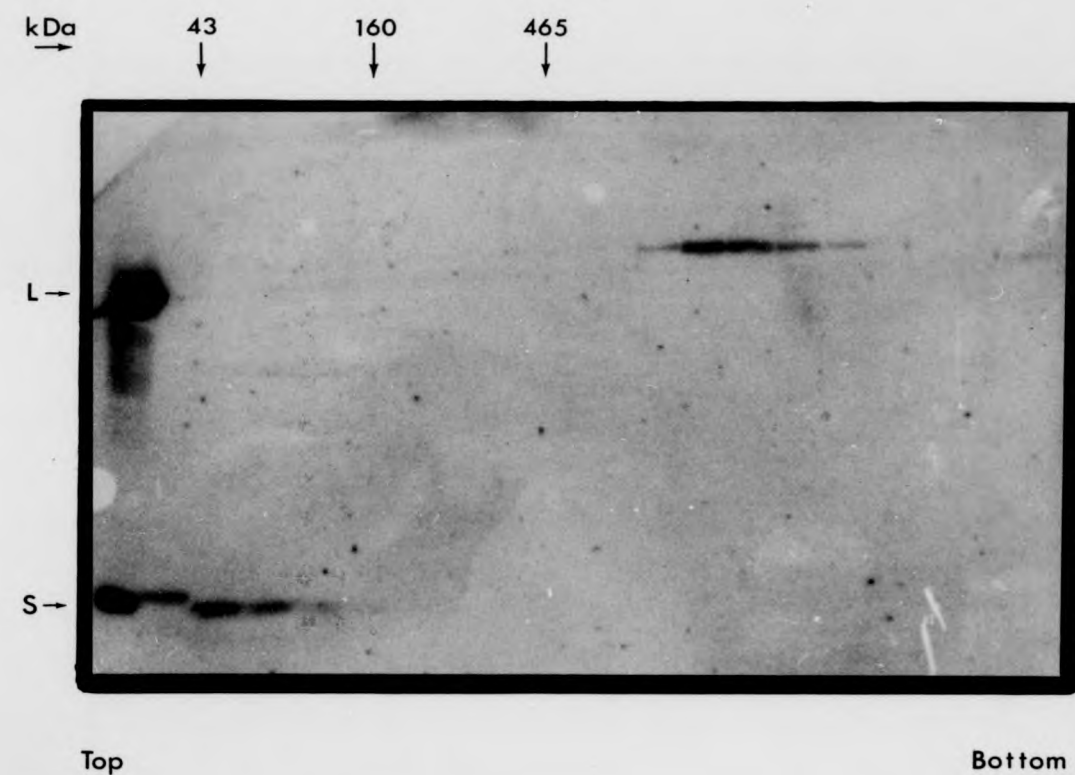
denaturing conditions (Fig. 68C, lane 4). When both the wheat Rubisco large and small subunits are expressed together in *E.coli* (Fig. 68C, lane 5) the antibodies against wheat Rubisco recognize polypeptides that comigrate with the *E.coli* chaperonin oligomer (Fig. 68B, lane 5). Whether these polypeptides are either large subunits or both large and small subunits is not known. It is possible that the binding of the large subunits to the chaperonin is a prerequisite for binding of the small subunits.

It is important to establish the form of the synthesized Rubisco small subunit in *E.coli*, since its association with the chloroplast chaperonin, but its possible failure to bind to the *E.coli* chaperonin may indicate why assembly of the wheat Rubisco enzyme in *E.coli* does not occur. The conformation of the wheat Rubisco mature small subunits when synthesized in the same *E.coli* cell as the large subunits is monomeric or dimeric (Bradley *et al.*, 1986); however this form could have been induced by the presence and/or association of the large subunits with the *E.coli* chaperonin. To examine the conformation of synthesized Rubisco small subunits in *E.coli*, an *E.coli* extract of TG2 cells containing pDB85 was analyzed by sucrose density gradient centrifugation, followed by analysis of the fractions on an SDS polyacrylamide gel which was subsequently immunoblotted. Most antibody preparations contain antibodies to the bacterial common antigen, as well as to the antigen that was injected (Section 2.2.13). In this experiment advantage was taken of this occurrence by using a non-purified antibody preparation, against wheat Rubisco enzyme which contains antibodies that will recognize the *E.coli* chaperonin polypeptides as well as antibodies that will cross-react with the wheat Rubisco large and small subunits. The Rubisco small subunits synthesized in *E.coli* were detected in the upper fractions of the sucrose gradient indicating a monomeric and/or dimeric form (Fig. 69). The *E.coli* chaperonin was detected as an oligomer with a  $M_r > 465000$ ; however no Rubisco small subunits were detected in fractions that contain the *E.coli* chaperonin oligomer (Fig. 69).

It is concluded that the *E.coli* chaperonin assembles into an oligomeric form and that this assembly is not dependent on the presence of the wheat chloroplast  $\alpha$

Figure 69. Analysis of *Triticum aestivum* Rubisco small subunits in *E.coli* extracts by sucrose density gradient centrifugation

A soluble protein extract in 100 mM Tris-HCl pH 8.0, containing 20 mM MgCl<sub>2</sub>, 10 mM NaHCO<sub>3</sub>, 1 mM EDTA, 1mM 2-mercaptoethanol and 1 mM PMSF, of *E.coli* TG2 cells containing pBD85 was fractionated on a sucrose density gradient [7-25% (w/v)] and the fractions analyzed as described in the legend to Figure 64. A crude antibody preparation against the wheat Rubisco holoenzyme was used. A wheat leaf extract was used as a marker for the Rubisco large (L) and small (S) subunits. Calibration markers for the sucrose gradient are indicated at the top, in terms of kDa.





chaperonin (cf Fig. 65A and Fig. 69). Wheat Rubisco large subunits synthesized in *E.coli* are associated with the *E.coli* chaperonin oligomer, while the small subunits do not bind either to the *E.coli* chaperonin or to the chaperonin-large subunit complex. In isolated pea chloroplasts the situation is different as the imported Rubisco small subunits bind to the chloroplast chaperonin, probably prior to their assembly into the holoenzyme.

#### 3.4.4 CO-EXPRESSION OF THE *TRITICUM AESTIVUM* CHLOROPLAST ALPHA CHAPERONIN AND THE RUBISCO LARGE AND SMALL SUBUNITS IN THE SAME *ESCHERICHIA.COLI* CELL

##### 3.4.4.1 Construction of the expression vector pSV16

The DNA sequences encoding the wheat Rubisco large and mature small subunits, and the chloroplast  $\alpha$  chaperonin are present in three different expression vectors pSV15 and, M68 and pDB85 respectively (Fig. 61; Fig. 67). The expression vectors pDB85 and pSV15 are derivatives of the pUC plasmids pUC18 and pUC9 respectively (Vieira & Messing, 1982; Norrander *et al.*, 1983) and are thus not compatible. To be able to express all three wheat sequences in the same *E.coli* cell either the  $\alpha$  chaperonin or the Rubisco small subunit sequence has to be transferred to a different compatible plasmid. An alternative possibility is to transfer the small subunit sequence to the expression vector containing the chloroplast  $\alpha$  chaperonin sequence with the advantage that these two sequences will be present in identical copy number in the *E.coli* cell.

To generate such an expression vector a fragment of about 2 kb containing the entire wheat Rubisco small subunit sequence was isolated from pDB85 by digestion with the restriction enzymes *Bgl*II and *Hind*III, following by the end-filling of the created recessed 3' ends. This blunt ended fragment was subsequently ligated into the plasmid pSV15 which had been linearized with *Eco*RI and the recessed 3' ends filled.



The plasmid in which the transcriptional direction for the  $\alpha$  chaperonin and the small subunit is identical was named pSV16 (Fig. 70), while the construct containing the small subunit fragment in the opposite orientation was called pSV17 (data not shown). The plasmid pSV16 was used for expression of the  $\alpha$  chaperonin and Rubisco small subunit polypeptides in *E.coli* (Section 3.4.43).

#### 3.4.4.2 Synthesis of the *Triticum aestivum* chloroplast $\alpha$ chaperonin and the Rubisco large and small subunits in the same *Escherichia coli* cell

The Rubisco large and small subunit sequences and the chloroplast  $\alpha$  chaperonin sequence in the two different constructs (pSV16 and M68) are all under the control of the *Plac* promoter (Fig. 67; Fig. 70). To examine the synthesis of the different polypeptides in the same *E.coli* cell, soluble protein extracts of *E.coli* TG2 cells containing pSV16, and of TG2 cells containing pSV16 that had been infected with M68, were analyzed by SDS polyacrylamide gel electrophoresis, followed by immunoblotting and detection of the polypeptides using different antibody preparations against the chaperonin  $\alpha$  subunit and the Rubisco large and small subunits. The chaperonin  $\alpha$  subunit encoded by pSV16 was detected in both the *E.coli* extracts (Fig. 71A). The Rubisco small subunit polypeptide was detected in cell extracts from TG2 cells containing pSV16 (Fig. 71B), while Rubisco large subunits were detected in extracts of cells that had been infected with M68 (Fig. 71B, lane 2). The faint band with a Mr of about 60 kDa observed when the Rubisco antibody preparation was used (Fig. 71B) is probably the result of the cross-reactivity of the *E.coli* chaperonin with antibodies against the bacterial common antigen which are present in the Rubisco antibody preparation. It is concluded that the Rubisco large and small subunits, and the chloroplast  $\alpha$  chaperonin are synthesized as stable soluble proteins in the same *E.coli* cell.

Figure 70. Construction of the expression plasmid pSV16

To construct pSV16 the *Bgl*I - *Hind*III fragment was isolated from pDB85; this fragment containing the wheat Rubisco small subunit fusion protein sequence, was ligated as a blunt end fragment into pSV15 that had been cleaved with *Eco*RI followed by filling of the recessed 3' ends. transcription of both the wheat sequences occurs from the lac promoter in the direction indicated by the arrows. The positions of the *Bgl*I (B), *Hind*III (H), *Eco*RI (E), the blunt end filled restriction sites (\*) and the ampicillin gene (amp) are indicated.

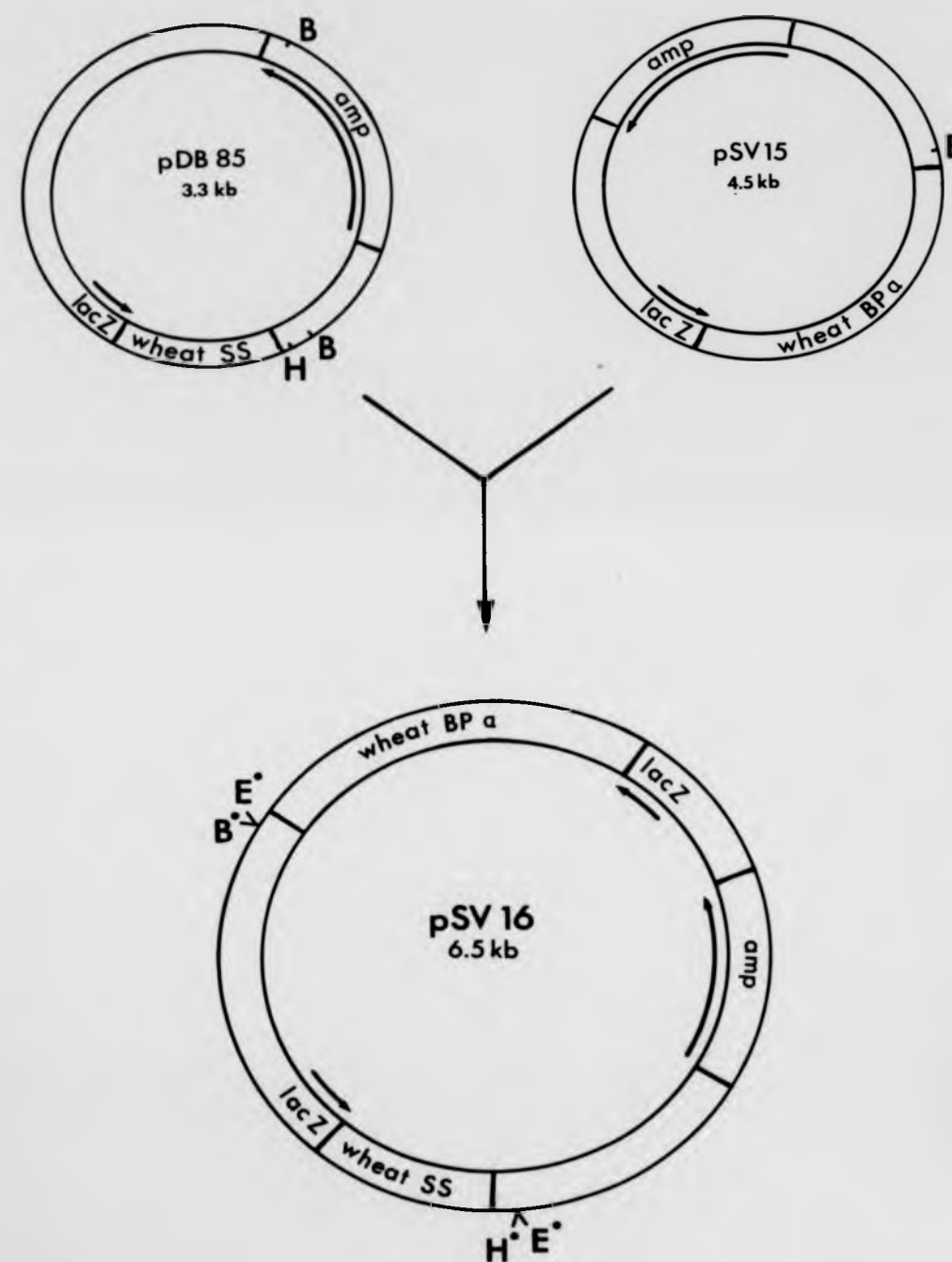
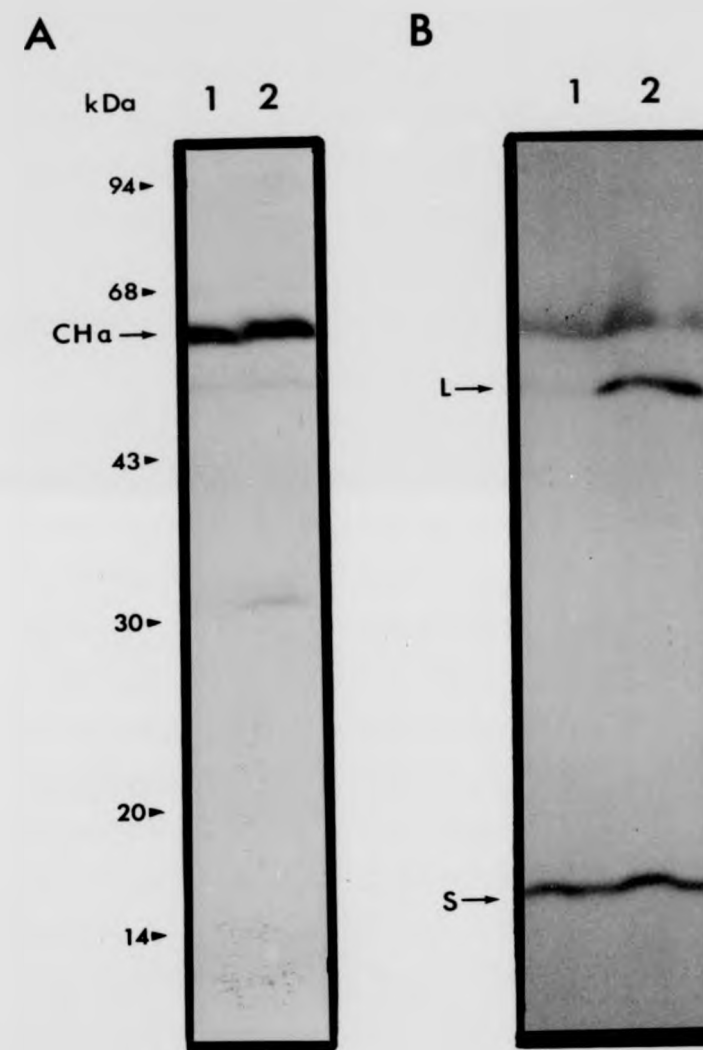


Figure 71. Analysis of the *Triticum aestivum* chloroplast  $\alpha$  chaperonin and the Rubisco large and small subunits in the same *E.coli* cell

Soluble protein extracts in 100 mM Tris-HCl pH 8.0, containing 20 mM MgCl<sub>2</sub>, 10 mM NaHCO<sub>3</sub>, 1 mM EDTA, 1 mM 2-mercaptoethanol and 1 mM PMSF of *E.coli* TG2 cells containing pSV 16 (lane 1), and *E.coli* cells containing pSV16 and infected with M68 (lane 2), were analyzed on a 15% (w/v) polyacrylamide gel containing SDS, followed by immunoblotting using either chloroplast chaperonin specific antibodies (Panel A) or specific antibodies against the Rubisco holoenzyme (Panel B). The positions of the chloroplast  $\alpha$  chaperonin (CHa) and the Rubisco large (L) and small (S) subunits are indicated.



No enzymic Rubisco activity was detected in an *E.coli* extract that contained all the three proteins (data not shown). To determine the form of both the Rubisco subunits in *E.coli* extracts containing the chloroplast  $\alpha$  chaperonin, analysis under non-denaturing conditions is required but time did not permit this. The failure to assemble an active plant Rubisco in *E.coli* in the presence of the chloroplast  $\alpha$  chaperonin indicates the complexity of the assembly pathway in plant chloroplasts, and suggests that other additional plant factors are probably required.

#### 3.4.5 DISCUSSION

The *E.coli* chaperonin exists in different oligomeric forms in *E.coli* extracts, as judged by analysis on non-denaturing gels. A low mobility form, probably the assembled 14-subunit oligomer is the major form of the *E.coli* chaperonin. However other forms, possibly either monomers or dimers, are also detectable (Fig. 64). A similar observation has been made for the chloroplast chaperonin in stromal extracts of *Pisum sativum* chloroplasts (Musgrove & Ellis, 1986; Musgrove *et al.*, 1987). In such extracts, the chloroplast chaperonin undergoes a reversible dissociation and MgATP causes the equilibrium of this dissociation process to shift towards the monomeric subunits (Musgrove & Ellis, 1986). Both the purified *E.coli* and chloroplast chaperonins show weak ATPase activity and dissociate in the presence of MgATP. Whether the dissociation of the purified chaperonins is reversible is unknown, but this seems unlikely as additional dialysable stromal factors have been suggested to be required for the reversibility (Musgrove & Ellis, 1986).

The oligomeric form of the *E.coli* chaperonin is observed when *E.coli* extracts are analyzed by sucrose density gradient centrifugation in the presence of MgCl<sub>2</sub> but not in its absence (cf. Fig. 64A and Fig. 65A). In an extract of *E.coli* cells containing the chloroplast chaperonin, heterologous chaperonin complexes are detected (Fig. 64). The chaperonin oligomer ( $M_r > 465\ 000$ ) detected in such an extract probably consists of both the *E.coli* and wheat chloroplast  $\alpha$  chaperonins and is only observed in the

presence of  $MgCl_2$  (Fig. 63; cf. Fig. 64B and Fig. 65B). These observations suggest that the effect of  $Mg^{2+}$  ions is identical for homologous and heterologous chaperonin complexes. The  $Mg^{2+}$  ions may interact with the chaperonin monomeric subunits thereby shifting the dissociation-association equilibrium towards the oligomeric form. However the interaction of  $Mg^{2+}$  ions with other proteins present in the lysed cell extract could prevent inhibition of chaperonin subunit assembly. It is not possible to distinguish between these possibilities at the moment.

The detection in *E.coli* extracts of heterologous protein complexes consisting of *E.coli* and wheat chloroplast  $\alpha$  chaperonin subunits is not surprising, as it is thought that the chloroplast  $\alpha$  chaperonin subunits are assembled with equal amounts of  $\beta$  chaperonin subunits in stromal extracts of *P.sativum* (Musgrove & Ellis, 1986). The amino acid sequences of the chloroplast  $\alpha$  and  $\beta$  chaperonins are about 50% identical, very similar to the amino acid sequence identity of the wheat chloroplast  $\alpha$  and *E.coli* chaperonins, i.e. 46% (Table 15). Thus the chloroplast  $\alpha$  chaperonin may accept the *E.coli* chaperonin as its 50% identical partner, resulting in assembly of heterologous rather than homologous chaperonin complexes. Further studies are needed to establish whether these hybrid chaperonins are functionally active.

Newly-synthesized Rubisco large subunits are found associated with the chaperonin oligomer in the stromal fraction of *Pisum sativum* chloroplasts (Barracough & Ellis, 1980), while *Triticum aestivum* Rubisco large subunits synthesized in *E.coli* are associated with the *E.coli* chaperonin oligomer but not with the higher mobility form (Fig. 68). When Rubisco small subunits are imported into chloroplasts *in vitro*, they rapidly assemble into the holoenzyme but a proportion is associated with the chloroplast chaperonin (Fig. 66). The rate of assembly of the imported small subunits is faster than the rate of assembly of newly-synthesized large subunits in isolated chloroplasts (Fig. 66; Ellis & van der Vies, 1988; Gatenby *et al.*, 1988). Because isolated chloroplasts are deprived of a source of imported small subunits the rate of Rubisco assembly may be determined by the available pool of small subunits. Schmidt and Mishkind (1983) reported that unassembled small

subunits are unstable in chloroplasts; the degradation of unassembled small subunits will reduce the poolsize even further.

When *Triticum aestivum* Rubisco mature small subunits are synthesized in *E.coli* they are not associated with the *E.coli* chaperonin oligomer but form stable monomers and possible dimers (Fig. 69). This failure to bind to the *E.coli* chaperonin oligomer is not the result of the absence of wheat Rubisco large subunits since coexpression of both the Rubisco subunits in *E.coli* does not alter the form of the mature small subunits (Bradley *et al.*, 1986).

Processing of small subunit precursors to mature small subunits within the chloroplast is known to occur in two steps (Bedbrook *et al.*, 1980; Robinson & Ellis, 1984; Mishkind *et al.*, 1984), generating an intermediate that retains a carboxyl-terminal part of the presequence. The significance of this two-step processing is not known; however it seems possible in principle that the carboxyterminal part of the presequence contains information required for the recognition and binding to the chloroplasts chaperonin-large subunit complex and thus is essential for Rubisco assembly. This idea is strengthened by the observation that anti-idiotypic antibodies against the small subunit presequence cross-react with the large-subunits (Pain *et al.*, 1988), while imported foreign proteins that are fused to the small subunit presequence are all found associated with the chloroplast chaperonin (Lubben *et al.*, 1989). If this idea is correct, small subunit precursor polypeptides made in *E.coli* may possibly associate with the *E.coli* chaperonin and permit Rubisco assembly these possibilities should be tested. However the synthesis of stable small subunit precursor polypeptides in *E.coli* appears to be a problem, since rapid processing to the mature subunit occurs (Landry & Bartlett, 1989), although in these experiments large subunits were not present. Both the presequence and regions within the mature polypeptide of the outer membrane protein A (omp A) of *E.coli* participate in the recognition of the purified *E.coli* chaperonin *in vitro* (Lecker *et al.*, 1989), suggesting that additional information present in the presequence may be needed for recognition and/or enhanced association to chaperonins in organelles such as chloroplasts and mitochondria.

The wheat chloroplast  $\alpha$  chaperonin and the wheat Rubisco large and small subunits are synthesized as stable soluble polypeptides in the same *E.coli* cell (Fig. 71). However it is unlikely that correct assembly into an oligomeric enzyme has occurred since no Rubisco activity was detected in an *E.coli* extract that contains all three polypeptides. Further investigations are required to determine the form of the wheat Rubisco large and small subunits in the presence of the chloroplast  $\alpha$  chaperonin in *E.coli*.

The failure of the chloroplast Rubisco subunits expressed in *E.coli* in the presence of the chloroplast  $\alpha$  chaperonin to assemble into an active enzyme may result from a number of different problems:

1. It may be necessary for the chloroplast  $\beta$  chaperonin to be present in addition to the  $\alpha$  chaperonin; this idea is supported by the presence of  $\alpha$  and  $\beta$  subunits in equal amounts in the chaperonin oligomer purified from chloroplasts (Musgrove & Ellis, 1986).
2. It may be that the information essential for assembly of a plant Rubisco enzyme is present in the presequence of the Rubisco small subunit precursor as described above; in contrast is the assembly of mature wheat Rubisco small subunits both with *A.nidulans* large subunits in *E.coli* (van der Vies *et al.*, 1986) and with *A.nidulans* large subunit cores *in vitro* (Andrews & Ballment, 1983) to produce an active Rubisco enzyme.
3. It may be that a chloroplast homologue to the *groES* protein of *E.coli* exists which is also required for Rubisco assembly; evidence in favour of this idea comes from recent work by Goloubinoff *et al* (1989) who showed that the assembly of *R. rubrum* large subunit dimers and the *A.nidulans* Rubisco oligomer in *E.coli* requires the activities of both the *E.coli* chaperonin and the *groES* protein. However as yet there is no published evidence for the occurrence of *groES*-like proteins in eukaryotic cells.

All these possibilities can be tested in principle, and current experiments in collaboration with S.M. Hemmingsen aim to express the  $\alpha$  and  $\beta$  subunits of the



*Brassica napus* chloroplast chaperonin in the same *E.coli* cell as the wheat Rubisco large and small subunits in the hope of rescuing Rubisco assembly.

#### 4 CONCLUDING REMARKS

---

---

#### 4 CONCLUDING REMARKS

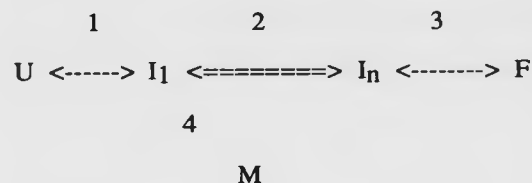
It has commonly been assumed that the folding and assembly of proteins in cells occurs spontaneously without the assistance of additional factors, and that the formation both necessary and sufficient to fold a polypeptide chain into a biologically functional three-dimensional structure resides within the amino acid sequence. However all the evidence for this notion of self-assembly comes from studies on the refolding of purified denatured proteins *in vitro*, and thus it has always remained formally possible that protein folding *in vivo* is assisted (Ellis & Hemmingsen, 1989). The work described in this thesis supports the view that assisted protein folding and assembly is likely to occur *in vivo*. During the course of this work a rapidly increasing number of papers has been published supporting the idea that proteins called molecular chaperones assist protein folding and assembly, as proposed by Ellis (1987). It is outside the scope of this thesis to discuss the family of molecular chaperones in detail. Whereas some members have already been discussed, the currently recognized molecular chaperones and their proposed roles are listed in Table 21 (Ellis, van der Vies & Alldrick, 1989).

Clearly the most interesting question arisen from this thesis is by what mechanisms these molecular chaperones function; what does each molecular chaperone recognize in a range of unrelated proteins and what is the underlying mechanism by which they catalyse protein folding? Studies on the refolding of denatured proteins *in vitro* have shown that partly folded intermediates can be isolated, suggesting that folding *in vitro*, and hence possibly *in vivo* occurs in a series of steps. However in these studies not all reactions lead to the formation of biologically functional i.e. correct structures, as the production of non-functional aggregates is often observed *in vitro*. The pathway of such a refolding process may thus be written as follows:-

Table 21. Molecular Chaperones

Class	Proposed roles	References
Nucleoplasmins	Nucleosome assembly	Laskey <i>et al.</i> (1978)
Chaperonins	Protein folding Oligomer assembly Protein transport DNA replication ? mRNA turnover ?	Hemmingsen <i>et al.</i> (1988); Ellis <i>et al.</i> (1989)
Bip-hsp 70	Protein transport Oligomer assembly Oligomer disassembly DNA replication	Pelham (1986; 1988); Deshaies <i>et al.</i> (1988); Chirico <i>et al.</i> (1988); Rothman and Schmid (1986); Zylic <i>et al.</i> (1989)
Signal recognition particle (SRP)	Protein transport	Walter and Lingappa (1986)
SecB protein	Protein transport	Collier <i>et al.</i> (1988)
Trigger factor	Protein transport	Crooke and Wickner (1987)
Pro-sequence	Subtilisin assembly	Zhu <i>et al.</i> (1989)
Ubiquitin	Ribosomal assembly	Finley <i>et al.</i> (1989)

See also Ellis, van der Vies & Alldrick (1989)



in which : U = denatured unfolded polypeptide

$I_1 \xrightarrow{\hspace{1cm}} I_n$  = partly folded intermediates

F = correctly folded polypeptide

M = misfolded polypeptide (aggregates)

It has been proposed that molecular chaperones prevent the formation of misfolded proteins by inhibiting reaction(s) 4. They achieve this by binding to the partly folded intermediates and hence guide them along pathways that result in the formation of correct structures (Ellis & Hemmingsen, 1989; Ellis, van der Vies, Alldrick, 1989). Support for the idea that molecular chaperones prevent the formation of misfolded structures rather than disrupting them comes from observations that a number of molecular chaperones, including the chaperonin, trigger factor and *secB* bind only to either unfolded or partly folded polypeptide chains, but not to misfolded or correctly folded protein structures.

To study the role of molecular chaperones it will be necessary to demonstrate that a polypeptide chain can be correctly refolded *in vitro* under physiological conditions only with the aid of a molecular chaperone. Earlier this year Goloubinoff *et al.* (1989) showed that the folding and assembly of both the Rubisco dimeric and hexadecameric enzymes in *E. coli* depends on both the endogenous *groEL* (large) and *groES* (small) chaperonins. Recently they have demonstrated in a series of elegant experiments that the dimeric form of Rubisco can be reconstituted from an unfolded biologically inactive polypeptide to a catalytically active dimer under physiological conditions only in the presence of the large and small chaperonins and MgATP (P. Goloubinoff, J.T. Christeller, A.A. Gatenby & G.H. Lorimer, submitted for publications). The reconstitution of Rubisco is an ordered process, which is observed only when the unfolded/partly folded polypeptide is allowed to bind to the large

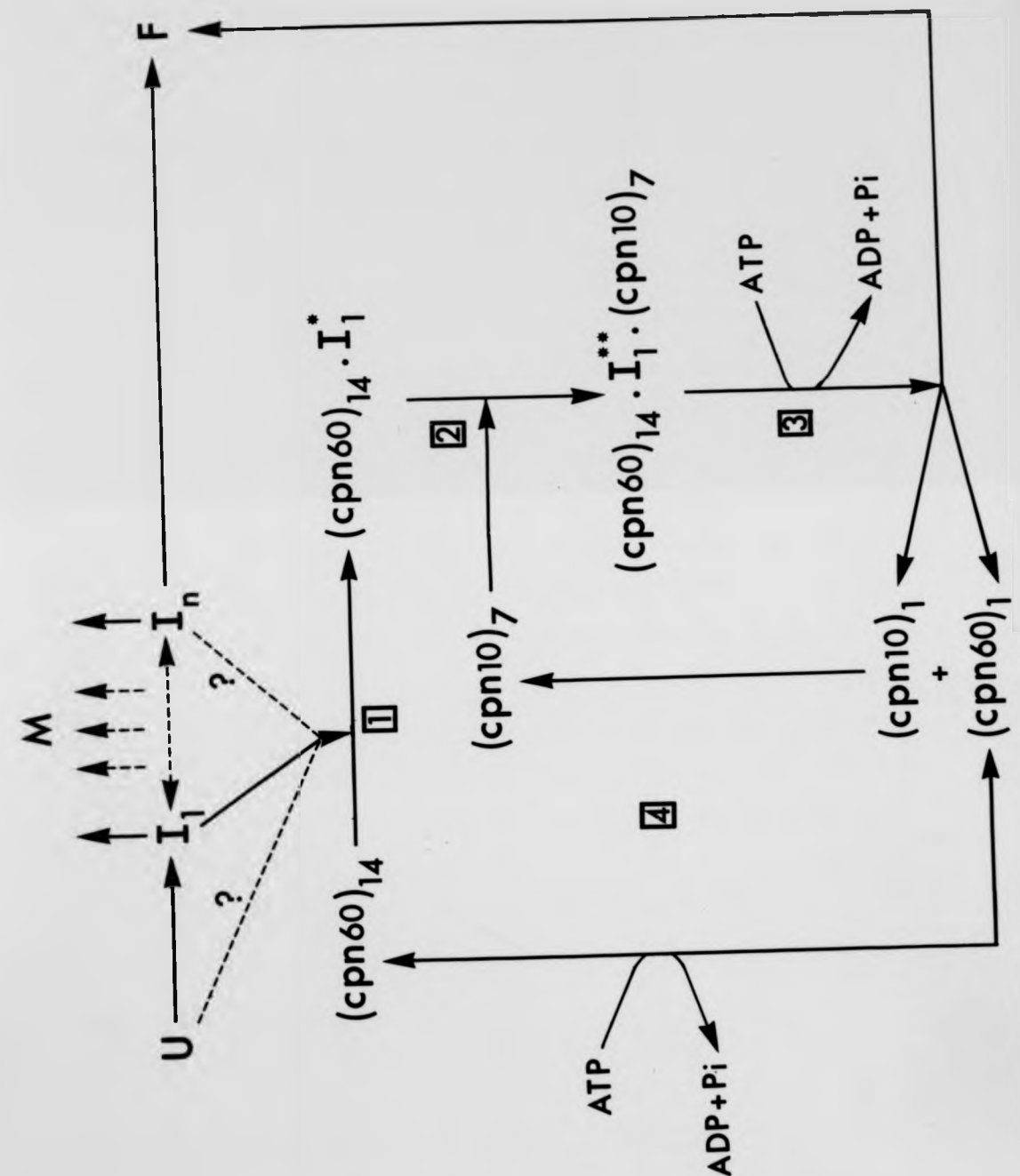
chaperonin oligomer to form a binary complex prior to association with the small chaperonin. A model for the folding pathway based on these studies is presented in Figure 72. This recently achieved reconstitution of Rubisco *in vitro* will allow studies as to what regions are involved in binding within both the chaperonins as well as within the Rubisco polypeptides.

The precise role of ATP in chaperonin action is unknown, but the observation that the *E.coli* large chaperonin oligomer contains 14 ATP-binding sites per oligomer (A.Girshovich, personal communication) suggests that probably each folded chaperonin monomer possesses one ATP-binding site. This estimate would be in agreement with the presence of the putative reversed dinucleotide binding site sequence identified in the large chaperonin subunit described in this thesis. The identification of this putative reversed dinucleotide binding site is a most remarkable observation since if it turns out to be correct it will have consequences for the identification of binding sites in general, for this observation suggests that it may be possible to create a particular binding site from a given set of amino acids present in more than one order along the polypeptide chain. It is not surprising that reversed nucleotide binding sites have not been identified previously since database searches are normally performed with sequences running from the amino to the carboxyterminus.

It is clearly important to establish whether the identified putative ATP-binding site actually binds ATP, and this can be determined by means of photo-crosslinking with radiolabelled ATP, followed by peptide mapping and determination of the amino acid sequences of the radiolabelled peptide. To determine whether a given set of amino acids can be folded in a particular fashion to create, e.g. a dinucleotide binding site, computer modelling of protein domains will be required. In order to establish whether this phenomenon has a more general occurrence it will be necessary to analyse reversed sequences of other known binding sites and to search databases with reversed sequences. These studies are currently in progress in collaboration with R. Hubbard and G.G. Dodson of the Department of Chemistry at York University.

Figure 72. A proposed model for chaperonin activity

A schematic diagram of assisted protein folding is shown based on observations made by Goloubinoff *et al.* (submitted for publication). Symbols: U; denatured/unfolded/newly-synthesized polypeptide;  $I_1$ ,  $I_1^*$ ,  $I_1^{**}$  and  $I_n$ , partly folded intermediates; F, correctly folded polypeptide; M, misfolded polypeptides (aggregates); cpn60, large chaperonin of *E. coli*; cpn10, small chaperonin of *E. coli*. The pathway of folding is proposed to occur in a series of reactions: 1, association of either U or I with the large chaperonin oligomer; 2, association of the small subunit chaperonin oligomer with the generated binary complex; 3, dissociation of the chaperonin-proteins complex by MgATP into chaperonin monomers and the release of the folded polypeptide; 4, formation of large chaperonin oligomers and small chaperonin oligomers.





## 5 REFERENCES

## 5 REFERENCES

- Amersham's Protocol (1985). Membrane transfer and detection methods. Amersham International.
- Amiri, J. Salnikow, J. & Vater, J. (1984). Amino-acid sequence of the large subunit of D-ribulosebiphosphate carboxylase/oxygenase from *Nicotiana tabacum*. Biochim. Biophys. Acta 784, 116-123.
- Anderson, C.W., Straus, J.W. & Dudock, B.S. (1983). Preparation of a cell-free protein-synthesizing system from wheat germ. Meth. of Enzymol. 101, 635-644.
- Andersson, I., Knight, S., Schneider, G., Lindqvist, Y., Linqvist, T, Branden, C-I & Lorimer, G.H. (1989). Crystal stucture of the active site of ribulose-bisphosphate carboxylase. Nature 337, 229-234..
- Andrews, T.J. & Ballment, B.J. (1983). The function of the small subunit of Ribulose bisphosphate carboxylase-oxygenase. J. Biol. Chem. 258, 7514-7518.
- Andrews. T.J. & Lorimer, G.H. (1985). Catalytic properties of a hybrid between cyanobacterial large subunits and higher plant small subunits of ribulose bisphosphate carboxylase-oxygenase. J. Biol. Chem. 260, 4632-4636.
- Andrews, T.J., Greenwood, D.M. & Yellowlees, D. (1984). Catalytically active hybrids formed *in vitro* between large and small subunits of different prokaryotic ribulose bisphosphate carboxylases. Arch. Biochem. Biophys. 234, 313-317.
- Andrews, T.J., Badger, M.R. & Lorimer, G.H. (1975). Factors affecting interconversion of kinetic form of ribulose diphosphate carboxylase-oxygenase from spinach. Arch. Biochem. Biophys. 171, 98-103.
- Ansorge, W. & de Mayer, L. (1980). Thermally stabilized very thin (0.02-0.03) polyacrylamide gels for electrophoresis. J. Chromatog. 202, 45-53.
- Arnon, D.I. (1949). Copper enzymes in isolated chloroplasts. Plant Physiol. 24, 1-15.
- Aubert, J.P., Biserte, G. & Loucheux-Lefebure, M.H. (1976). Carbohydrate-peptide linkage in glycoproteins. Arch. Biochem. Biophys. 175, 410-418.
- Badger, M.R., Kaplan, A. & Berry, J.A. (1980). Internal inorganic carbon pool of *Chlamydomonas reinhardtii*. Evidence for a carbon dioxide-concentrating mechanism. Plant Physiol. 66, 407-413.

- Bahr, J.T. & Jensen, R.G. (1974). Ribulose diphosphate carboxylase from freshly ruptured spinach chloroplasts having an *in vivo*  $K_m(\text{CO}_2)$ . *Plant Physiol.* 53, 39-44.
- Bankier, A.T. & Bottell, B.G. (1982). Shotgun DNA sequencing. Laboratory manual, Laboratory of Molecular Biology, MRC Centre, Cambridge, U.K.
- Barlow, D.J., Edwards, M.S. & Thornton, J.M. (1986). Continuous and discontinuous protein antigenic determinants. *Nature* 322, 747-748.
- Barracough, R. & Ellis, R.J. (1979). The biosynthesis of ribulose biphosphate carboxylase. Uncoupling of the synthesis of the large and small subunits in isolated soybean leaf cells. *Eur. J. Biochem.* 94, 165-177.
- Barracough, R. & Ellis, R.J. (1980). Protein synthesis in chloroplasts IX. Assembly of newly-synthesized large subunits into ribulose biphosphate carboxylase in intact pea chloroplasts. *Biochim. Biophys. Acta* 608, 19-31.
- Bedbrook, J.R., Coen, D.M., Beaton, A.R., Bogorad, L. & Rich, A. (1979). Location of the single gene for the large subunit of ribulose biphosphate carboxylase on the maize chloroplast chromosome. *J. Biol. Chem.* 254, 905-910.
- Bedbrook, J.R., Smith, S.M. & Ellis, R.J. (1980). Molecular cloning and sequencing of cDNA encoding the precursor to the small subunit of chloroplast ribulose-1,5-bisphosphate carboxylase. *Nature* 287, 692-696.
- Benedict, C.R. (1973). The presence of ribulose 1.5-diphosphate carboxylase in the non-photosynthetic endosperm of germinating castor beans. *Plant Physiol.* 51, 755-759.
- Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980). The ovalbumin gene-sequence of putative control regions. *Nucl. Acids Res.* 8, 127-142.
- Berget, S. (1984). Are U4 small nuclear ribonucleoproteins involved in polyadenylation ? *Nature* 309, 179-181.
- Berry, J.D., Nikolau, B.J., Carr, J.P. & Klessig, D.F. (1986). Translational regulation of light induced ribulose-1.5-bisphosphate carboxylase gene expression in *Amaranth*. *Mol. Cell. Biol.* 6, 2347-2353.

- Berry-lowe, S.L. McKnight, T.D., Shah, D.M. & Meagher, R.B. (1982). The nucleotide sequence, expression, and evolution of one member of a multigene family encoding the small subunit of ribulose-1.5.-bisphosphate carboxylase in soybean. *J. Mol. Appl. Genet.* **1**, 483-498.
- Biggin, M.D., Gibson, T.J. & Hong, G.F. (1983). Buffer gradient gels and <sup>35</sup>S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**, 3963-3965.
- Birktoft, J.J., Fernley, R.T., Bradshaw, R.A. & Banaszak, L.J. (1982). Amino acid sequence homology among the 2-hydroxy acid dehydrogenases: Mitochondrial and cytoplasmic malate dehydrogenases form a homologous system with lactate dehydrogenase. *Proc. Natl. Acad. Sci. USA* **79**, 6166-6170.
- Birnboim, H.C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7**, 1513-1523.
- Birnstiel, M.L., Busslinger, M. & Strub, K. (1985). Transcription termination and 3' processing: The end is in site ! *Cell* **41**, 349-359.
- Blair, G.E. & Ellis, R.J. (1973). Protein synthesis in chloroplasts. I. Light-driven synthesis of the large subunit of Fraction I protein by isolated chloroplasts. *Biochim. Biophys. Acta* **319**, 223-234.
- Bleasby, A., Akrigg, D. & Wootton, J. (1988). A non reductant, integrated composite sequence database with search and retrieval software. Integrated Sequences/Integrated Structure Data Resource, British Technology Group.
- Bloom, M.V., Milos, P. & Roy, H. (1983). Light-dependent assembly of ribulose-1,5-bisphosphate carboxylase. *Proc. Natl. Acad. Sci. USA* **80**, 1013-1017.
- Bochkareva, E.S., Lissin, N.M. & Girshovich, A.S. (1988). Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. *Nature* **336**, 254-257.
- Bole, D.G., Hendershot, L.M. & Kearney, J.F. (1986). Posttranslational association of immunoglobulin heavy chain binding protein with nascent heavy chains in nonsecreting and secreting hybridomas. *J. Cell Biol.* **102**, 1558-1566.
- Bonner, T.I., Brenner, D.J., Neufeld, B.R. and Britten, R.J. (1973). Reduction of the rate of DNA reassociation by sequence divergence. *J. Mol. Biol.* **81**, 123-135.
- Bovenberg, W.A., Kroes, R.E., Kool, A.J. & Nijkamp, H.J.J. (1984). Physical mapping, nucleotide sequencing and expression in *E.coli* minicells of the gene for the large subunit of ribulose bisphosphate carboxylase from *Petunia hybrida*. *Curr. Genet.* **8**, 231-241.

- Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R. & Burgess, J. (1986). Posttranslational processing of Concanavalin A precursors in jack bean cotyledons. *J. Cell Biol.* 102, 1284-1297.
- Boyle, S.A., Hemmingsen, S.M. & Dennis, D.T. (1986). Uptake and processing of the precursor to the small subunit of ribulose 1.5-bisphosphate carboxylase by leucoplasts from endosperm of developing castor oil seeds. *Plant Physiol.* 81, 817-822.
- Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, 72, 248-254.
- Bradley, D. & Gatenby, A.A. (1985). Mutational analysis of the maize chloroplast ATPase- subunit gene promoter mutants in *E.coli* and their characterization in a chloroplast *in vitro* transcription system. *EMBO J.* 4, 3641-3648.
- Bradley, D., van der Vies, S.M. & Gatenby, A.A. (1986). Expression of cyanobacterial and higher-plant ribulose 1,5-bisphosphate carboxylase genes in *Escherichia coli*. *Phil. Trans. R. Soc. Lond.* B313, 447-458.
- Broglie, R. Coruzzi, G., Lamppa, G., Keith, B. & Chua, N-H. (1983). Structural analysis of nuclear genes encoding for the precursor of the small subunit of wheat ribulose-1.5-bisphosphate carboxylase-oxygenase. *Biochem. J.* 183, 747-750.
- Broglie, R. Coruzzi, G., Fraley, R.T., Rogers, S.G., Horsch, R.B., Niedermeyer, J.G., Fink, C.L., Flick, J.S. & Chua, N-H. (1984). Light-regulated expression of a pea ribulose-1.5-bisphosphate carboxylase small subunit gene in transformed plant cells. *Science* 224, 838-843.
- Broglie, R. Coruzzi, G., Keith, B. & Chua, N-H. (1984). Molecular biology of C4 photosynthesis in *Zea mays*: differential location of proteins and mRNAs in two leaf cell types. *Plant Mol. Biol.* 3, 431-444.
- Buchanan, B.B. (1980). Role of light in the regulation of chloroplast enzymes. *Ann. Rev. Plant Physiol.* 31, 341-374
- Cannon, S., Wang, P. & Roy, H. (1986). Inhibition of ribulose biphosphate carboxylase assembly by antibody to a binding protein. *J. Cell Biol.* 103, 1327-1335.
- Cashmore, A.R. (1976). Protein synthesis in plant leaf tissue: The sites of synthesis of the major proteins. *J. Biol. Chem.* 251, 2848-2853

- Cashmore, A.R. (1983). Nuclear genes encoding the small subunit of ribulose-1.5-bisphosphate carboxylase. In Genetic engineering of plants (eds. T.Kosuge, C.P. Meedith & A. Holleander). Basic life sciences vol 26 Plenum press New York, London pp 29-38.
- Cashmore, A.R., Broadhurst, M.K. & Gray, R.E. (1978) Cell-free synthesis of leaf proteins: identification of an apparent precursor of the small subunit of ribulose-1.5-bisphosphate carboxylase Proc. Natl. Acad. Sci. USA 75, 655-659.
- Cashmore, A.R., Szabo, L. Timko, M. Kausch, A. van den Broeck, G., Schreier, P., Bohnert, . Herrera-Estrella, L. van Montagu, M & Schell, J (1985). Import of polypeptides into chloroplasts. Biotechnology 3, 803-808
- Cattolica, R.A. (1986). Chloroplast evolution in algae and land plants. Trends Ecol. Evol. 1, 64-67.
- Chaconas, G. & van de Sande, J.H. (1980). 5'-<sup>32</sup>P labeling of RNA and DNA restriction fragments. Meth. Enzymol. 65, 75-85.
- Chanda, P.K., Ono, M., Kuwano, M. & Kung, H-F (1985). Cloning, sequence analysis and expression of alteration of the mRNA stability gene (*ams*<sup>+</sup>) of *Escherichia coli*. J. Bact. 161, 446-449.
- Chandhari, P., Cannon, S., Hubbs, A. & Roy, H. (1987). Abstracts 16th annual UCLA symposium. J. Cell Biochem. suppl. 11B, 50.
- Chandrasekhar, G.N., Tilly, K., Woolford, C., Hendrix, R.W. & Georgopoulos, C.P. (1986). Purification and properties of the *groES* morphogenetic protein of *Echerichia coli*. J. Biol. Chem. 261, 12414-12419.
- Chang, C., Kokontis, J. & Liao, S. (1988). Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors. Proc. Natl. Acad. Sci. USA 85, 7211-7215.
- Chapman, M.S., Smith, W.W., Suh, S.W., Cascio, D., Howard, A., Hamlin, R., Xuong, N.-H. & Eisenberg, D. (1986). Structural studies of the Rubisco from tobacco. Phil. Trans. R. Soc. Lond. B 313, 367-378.
- Chapman, M.S., Suh, S.W., Cascio, D., Smith, W.W. & Eisenberg, D. (1987). Sliding-layer conformational change limited by the quaternary structure of plant Rubisco. Nature 329, 354-351.
- Chase, J.W. & Williams, K.R. (1986). Single-stranded DNA binding proteins required for DNA replication. Ann. Rev. Biochem. 55, 103-136.

- Cheng, M.Y., Hartl, F.-U., Martin, J., Pollock, R.A., Kalousek, F., Neupert, W., Hallberg, E.M., Hallberg, R.L. & Horwich, A.L. (1989). Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature* 337, 620-625.
- Cherbas, L., Schulz, R.A., Koehler, M.M.D., Savakis, C. & Cherbas, P. (1986). Structure of the *Eip28/29* gene, an ecdysone-inducible gene from *Drosophila*. *J. Mol. Biol.* 189, 617-631.
- Chin, D.J., Gil, G., Russell, D.W., Liscum, L., Luskey, K.L., Basu, S.K., Okayama, H., Berg, P., Goldstein, J.L. & Brown, M.S. (1984). Nucleotide sequence of 3-hydroxy-3-methyl-glutaryl coenzyme A reductase, a glycoprotein of endoplasmic reticulum. *Nature* 308, 613-617.
- Chin, D.T., Goff, S.A., Webster, T., Smith, T. & Goldberg, A.L. (1988). Sequence of the *lon* gene in *Escherichia coli*: a heat-shock gene which encodes the ATP-dependent protease La. *J. Biol. Chem.* 263, 11718-11728.
- Chirio, W.J., Waters, M.G. & Blobel, G. (1988). 70k heat shock related proteins translocation into microsomes. *Nature* 332, 805-810.
- Chou, R.Y. & Fasman, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. in Enzymol.* 47, 45-147.
- Christeller, J.T., Terzaghi, B.E., Hill, D.F. & Laing, W.A. (1985). Activity expressed from cloned *Anacystis nidulans* large and small subunit ribulose biphosphate carboxylase genes. *Plant Mol. Biol.* 5, 257-263.
- Chua, N-H., & Schmidt, G.W. (1978). Post-translational transport into intact chloroplasts of a precursor to the small subunit of ribulose-1.5-bisphosphate carboxylase. *Proc. Natl. Acad. Sci. USA* 75, 6110-6114.
- Cline, K., Werner-Washburne, M., Lubben, T.H. & Keegstra, K. (1985). Precursors to two nuclear-encoded chloroplast proteins bind to the outer envelope membrane before being imported into chloroplasts. *J. Biol. Chem.* 260, 3691-3696.
- Coen, D.M., Bedbrook, J.R., Bogorad, L. & Rich, A. (1977). Maize chloroplast DNA fragment encoding the large subunit of ribulose biphosphate carboxylase. *Proc. Natl. Acad. Sci. USA* 74, 5487-5491.
- Collier, D.N., Bankaitis, V.A., Weiss, J.B. & Bassford, Jr., P.J. (1988). The antifolding activity of SecB promotes the export of the *E.coli* maltose-binding protein. *Cell* 53, 273-283.



- Cookson, J., Baird, P.N., Hall, L.M.C. & Coates, A.R.N. (1989). Identification of two unknown reading frames in *Synechococcus* as homologues of the 10K and 65K antigen genes in *E.coli* and *Coxiella burnetii*. Nucl. Acid Res. (in press).
- Coruzzi, G. Broglie, R., Cashmore, A.R. & Chua, N-H. (1983). Nucleotide sequence of two pea cDNA clones encoding the small subunit of ribulose-1.5-bisphosphate carboxylase and the major chlorophyll *a/b*-binding thylakoid polypeptide. J. Biol. Chem. 258, 1399-1402.
- Coruzzi, G. Broglie, R., Edwards, C. & Chua, N-H. (1984). Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1.5-bisphosphate carboxylase. EMBO J. 3, 1671-1679.
- Cowing, D.W., Bardwell, J.C.A., Craig, E.A., Woolford, C., Hendrix, R.W. & Gross, C.A. (1985). Consensus sequence for *Escherichia coli* heat shock gene promoters. Proc. Natl. Acad. Sci. USA 82, 2679-2683.
- Cozens, A.L. & Walker, J.E. (1987). The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium *Synechococcus* 6301. J. Mol. Biol. 194, 359-383.
- Craven, G.R., Steers, Jr., E. & Anfinsen, C.B. (1965). Purification, composition and molecular weight of the  $\beta$ -galactosidase of *Escherichia coli* K12. J. Biol. Chem. 240, 2468-2477.
- Creighton, T.E. (1984). Proteins, structures and molecular principles. Publ.; W.H. Freeman and Company, New York.
- Criddle, R.S., Dann, B., Kleinkopf, G. & Huffaker, R.C. (1970). Differential synthesis of Rubisco subunit. Biochem. Biophys. Res. Commun. 41, 621-627.
- Crooke, E., & Wickner, W. (1987). Trigger factor: A soluble protein that folds pro-OmpA into a membrane-assembly-competent form. Proc. Natl. Acad. Sci. 84, 5216-5220.
- Crooke, E., Brundage, L., Rice, M. & Wickner, W. (1988). ProOmpA spontaneously folds in a membrane assembly competent state which trigger factor stabilizes. EMBO J. 7, 1831-1835.
- Dagert, M. & Ehrlich, S.D. (1979). Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. Gene 6, 23-28.
- Davidson, B.E., Sajò, M., Noller, H.F., Harris, J.I (1967). Amino-acid sequence of glyceraldehyde 3-phosphate dehydrogenase from lobster muscle. Nature 216, 1181-1185.

- Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P. & Bedbrook, J. (1985). Linkage and homology analysis divides the eight genes for the small subunit of petunia ribulose-1.5-bisphosphate carboxylase into three gene families. *Proc. Natl. Acad. Sci. USA* 82, 4964-4968.
- Deshaies, R.J., Koch, B.D., Werner-Washburne, M., Craig, E.A. & Schekman, R. (1988). A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. *Nature* 332, 800-805.
- Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* 12, 387-395.
- Dix, N. & Wootton, Y. (1988). LUPES Leeds University Protein Engineering Software. Software for pattern recognition and feature discrimination in sequence database searches, Integrated Sequence/Integrated Structure Data Resource, British Technology Group.
- Dobberstein, B., Blobel, G. & Chua, N-H. (1977). *In vitro* synthesis and processing of a putative precursor for the small subunit of ribulose-1.5-bisphosphate carboxylase of *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA* 74, 1082-1085.
- Donnelly, M.I., Stringer, C.D. & Hartman, F.C. (1983). Characterization of the activator site of *Rhodospirillum rubrum* ribulose bisphosphate carboxylase/oxygenase. *Biochemistry, Wash.* 22, 4346-4352.
- Doolittle, R.F. (1986). Of URFS and ORFS. A primer on how to analyse derived amino acid sequences. University Science Books, U.S.A.
- Dron, M., Rahiv, M. & Rochaix, J.D. (1982). Sequence of the chloroplast DNA region of *Chlamydomonas reinhardtii* containing the gene of the large subunit of ribulose bisphosphate carboxylase and the parts of its flanking genes. *J. Mol. Biol.* 162, 775-793.
- Eckes, P. Schell, J. & Willmitzer, L. (1985). Organ-specific expression of three leaf/stem specific cDNAs from potato is regulated by light and correlated with chloroplast development. *Mol. Gen. Genet.* 199, 216-224.
- Edwards, G. & Walker, D.A. (1983). C3, C4: Mechanisms, and cellular and environmental regulation of photosynthesis. Blackwell Scientific Publications
- Eilers, M. & Schatz, G. (1988). Protein unfolding and the energetics of protein translocation across biological membranes. *Cell* 52, 481-483.

- Eisenberg, D. Almassy, R.J., Janson, C.A., Chapman, M.S., Suh, S.W., Cascio, D. & Smith, W.W. (1987). Some evolutionary relationships of the primary biological catalysts glutamine synthetase and Rubisco. In Cold Spring Harbor Symposia on Quantitative Biology Vol L II, pp 483-490.
- Ellis, R.J. (1978). Chloroplast proteins and their synthesis. In Plant proteins (ed. G. Norton) London: Butterworths, pp 25-40
- Ellis, R.J. (1979). The most abundant protein in the world. Trends Biochem. Sci. 4, 241-244.
- Ellis, R.J. (1981). Chloroplast proteins: Synthesis, transport and assembly. Ann. Rev. Plant Physiol. 32, 111-137.
- Ellis, R.J. (1986). Photoregulation of plant gene expression. Bioscience Reports 6, 127-135.
- Ellis, R.J. (1987). Proteins as molecular chaperones. Nature 328, 378-379.
- Ellis, R.J. & Gatenby, A.A. (1984). Ribulose biphosphate carboxylase: properties and synthesis. In Genetic manipulation of plants and its application to agriculture (eds. P.J. Lea and G.R. Stewart ) Oxford University Press pp 41-60.
- Ellis, R.J. & Hartley, M.R. (1982). Preparation of higher plant chloroplasts active in protein and RNA synthesis. In: Methods of chloroplast molecular biology (eds. M. Edelman, R.B. Hallick & N.H. Chua), Elsevier Biomedical Press, pp. 169-188.
- Ellis, R.J. & Hemmingsen, S.M. (1989). Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. Trends in Biochem. Sci. 14, 339-343.
- Ellis, R.J. & van der Vies, S.M. (1988). The Rubisco subunit binding protein. Photosynth. Res. 16, 101-115.
- Ellis, R.J., van der Vies, S.M. & Hemmingsen, S.M. (1987). The Rubisco large subunit binding protein - a molecular chaperone ? In: Plant Molecular Biology (eds. Von Wettstein, D. & Chua, N.-H.) 33-40 (Nato ASI Series A, Life Sciences Vol. 140, Plenum Press New York and London, 1987).
- Ellis, R.J., van der Vies, S.M. & Alldrick, S. (1989). Molecular chaperones and chloroplast biogenesis. Proc. 8th Int. Congress Photosyn. (Stockholm) in press.
- Ellis, R.J., van der Vies, S.M. & Hemmingsen, S.M. (1989). The molecular chaperone concept. Biochem. Soc. Symp. 55, 145-153.

- Emmerich, F., Thole, J., van Embden, J. & Kaufmann, S.H.E. (1986). A recombinant 65 kilodalton protein from *Mycobacterium bovis* bacillus Calmette-guerin specifically stimulates human T4 clones reactive to mycobacterial antigens. *J. Exp. Med.* 163, 1024-1029.
- Engers, H.D., Abe, M., Bloom, B.R., Mehra, O., Britton, W., Buchanan, T.M., Khanolkar, S.K., Young, D.B., Closs, O., Gillis, T., Harboe, M., Ivanyi, J., Kolk, A.H.J. & Shepard, C.C. (1985). Results of a world health organization-sponsored workshop on monoclonal antibodies to *Mycobacterium leprae*. *Infect. Immun.* 48, 603-605.
- Erion, J.L., Tarnowski, J., Peacock, S., Caldwell, P., Redfield, B., Brot, N. & Weissbach, H. (1983). Synthesis of the large subunit of ribulose-1,5-bisphosphate carboxylase in an *in vitro* partially defined *E.coli* system. *Plant Mol. Biol.* 2, 279-290.
- Fayet, O., Louarn, J.-M., & Georgopoulos, C.P. (1986). Suppression of the *Escherichia coli* *dnaA46* mutation by amplification of the *groES* and *groEL* genes. *Mol. Gen. Genet.* 202, 435-445.
- Feierabend, J. & Schrader-Reichhardt, U. (1976). Biochemical differentiation of plastids and other organelles in rye leaves with a high-temperature induced deficiency of plastid ribosomes. *Planta* 129, 133-145.
- Feierabend, J. & Wildner, G. (1978). Formation of the small subunit in the absence of the large subunit of ribulose 1.5 bisphosphate carboxylase in 70S ribosome-deficient rye leaves. *Arch. Biochem. Biophys.* 186, 283-291.
- Findly, R.C., Gillies, R.J., Shulman, R.G. (1983). *In vivo* phosphorus-31 nuclear magnetic resonance reveals lowered ATP during heat shock of *Tetrahymena*. *Science* 219, 1223-1225.
- Finley, D., Bartel, B. & Varshavsky, A. (1989). The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosomal biogenesis. *Nature* 338, 394-401.
- Fitzgerald, M. & Shenk, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24, 251-260.
- Fluhr, R. Fromm, H. & Edelman, M. (1983). Clone bank of *Nicotiana tabacum* chloroplast DNA: mapping of the alpha, beta and epsilon subunit of the ATPase coupling factor, the large subunit of ribulose bisphosphate carboxylase, and the 32-kDal membrane protein. *Gene* 25, 271-280.
- Freedman, R.B. (1987). Folding into the right shape. *Nature* 329, 196-197.

- Freedman, R.B. (1989). Protein disulfide isomerase: multiple roles in the modification of nascent secretory proteins. *Cell* 57, 1069-1072.
- Fry, D.C., Kuby, S.A. & Mildvan, A.S. (1986). ATP-binding site of adenylate kinase: mechanistic implications of its homology with *ras*-encoded p21, F1-ATPase, and other nucleotide-binding proteins. *Proc. Natl. Acad. Sci. USA* 83, 907-911.
- Fujiki, Y., Fowler, S., Shio, H., Hubbard, A.L. & Lazarow, P.B. (1982). Polypeptide and phospholipid composition of the membrane of rat liver peroxisomes: comparison with endoplasmic reticulum and mitochondrial membranes. *J. Cell Biol.* 93, 103-110.
- Fuller, R.S. & Kornberg, A. (1983). Purified *dnaA* protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. *Proc. Natl. Acad. Sci. USA* 80, 5817-5821.
- Fuller, R.S., Bertsch, L.L., Dixon, N.E., Flynn, J.E. Jr., Kaguni, J.M., Low, R.L., Ogana, T. & Kornberg, A. (1983). Enzymes in the initiation of replication at the *E. coli* chromosomal origin. In *Mechanisms of DNA Replication and Recombination*, UCLA Symp. Mol. Cell Biol., New Series, Vol. 10, N. Cozzarelli, ed. (New York: Alan R. Liss, Inc.) pp. 275-288.
- Fuller, R.S., Funnell, B.E. & Kornberg, A. (1984). The *dnaA* protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell* 38, 889-900.
- Fuller, R.S., Kaguni, J.M. & Kornberg, A. (1981). Enzymatic replication of the origin of the *Escherichia coli* chromosome. *Proc. Natl. Acad. Sci. USA* 78, 7370-7374.
- Furutani, Y., Morimoto, Y., Shibahara, S., Noda, M., Takahashi, H., Hirose, T., Asai, M., Inayama, S., Hayashida, H., Miyata, T. & Numa, S. (1983). Cloning and sequence analysis of cDNA for ovine corticotropin-releasing factor precursor. *Nature* 301, 537-540.
- Gallagher, T.A. & Ellis, R.J. (1982). Light-stimulated transcription of genes for two chloroplast polypeptides in isolated pea leaf nuclei. *EMBO J.* 1, 1493-1498.
- Gallagher, T.F., Jenkins, G.I., Smith, S.M. & Ellis, R.J. (1984). Photoregulation of the nuclear and chloroplast genes for ribulose-1,5-bisphosphate carboxylase. In *Chloroplast biogenesis*. Society for Experimental Biology Seminar Series 21 (ed. R.J. Ellis) Cambridge University Press pp 303-319.
- Gatenby, A.A. (1984). The properties of the large subunit of maize ribulose bisphosphate carboxylase/oxygenase synthesised in *Escherichia coli*. *Eur. J. Biochem.* 144, 361-366.

- Gatenby, A.A. & Castleton, J.A. (1982). Amplification of maize ribulose biphosphate carboxylase large subunit synthesis in *E.coli* by transcriptional fusion with the lambda N operon. *Mol. Gen. Genet.* 185, 424-429.
- Gatenby, A.A. & Cuellar, R.E. (1985). Antitermination is required for readthrough transcription of the maize *rbc L* gene by a bacteriophage promoter in *Escherichia coli*. *Eur. J. Biochem.* 153, 355-359.
- Gatenby, A.A., Castleton, J.A. & Saul, M.W. (1981). Expression in *E.coli* of maize and wheat chloroplast genes for the large subunit of ribulose biphosphate carboxylase. *Nature* 291, 117-121.
- Gatenby, A.A., van der Vies, S.M. & Bradley, D. (1985). Assembly in *E.coli* of a functional multi-subunit ribulose biphosphate carboxylase from a blue-green alga. *Nature* 314, 617-620.
- Gatenby, A.A., van der Vies, S.M. & Rothstein, S.J. (1987). Co-expression of both the maize large and wheat small subunit genes of ribulose biphosphate carboxylase in *Escherichia coli*. *Eur. J. Biochem.* 168, 227-231.
- Gatenby, A.A., Lubben, T.H., Ahlquist, P. & Keegstra, K. (1988). Imported large subunits of ribulose biphosphate carboxylase/oxygenase, but not imported b-ATP synthase subunits, are assembled into holoenzyme in isolated chloroplasts. *EMBO J.* 7, 1307-1314.
- Georgopoulos, C.P. & Eisen, H. (1974). Bacterial mutants which block phage assembly. *J. Supramol. Struct.* 2, 349-359.
- Georgopoulos, C.P., Hendrix, R.W. & Kaiser, A.D. (1972). Role of the host cell in bacteriophage morphogenesis: effect of a bacterial mutation on T4 head assembly. *Nature new Biol.* 239., 38-41.
- Georgopoulos, C.P., Hendrix, R.W., Casjens, S.R. & Kaiser, A.D. (1973). Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* 76, 45-60.
- Gething, M-J., McCammon, K. & Sambrook, J. (1986). Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport. *Cell*, 46, 939-950.
- Gibson, J.L. & Tabita, F.R. (1977). Different molecular forms of D-ribulose-1.5-bisphosphate carboxylase from *Rhodospseudomonas sphaeroides*. *J. Biol. Chem.* 252, 943-949.
- Gibson, J.L. & Tabita, F.R. (1986). Isolation of the *Rhodospseudomonas sphaeroides* form I ribulose-1.5-bisphosphate carboxylase/oxygenase large and small subunit genes and expression of the active hexadecameric enzyme in *Escherichia coli*. *Gene* 44, 271-278.

- Gillis, T.P., Miller, R.A., Young, D.B., Khanolkar, S.R. & Buchanan, T.M. (1985). Immunochemical characterization of a protein associated with *Mycobacterium leprae* cell wall. *Infect. Immun.* 49, 371-377.
- Goff, S.A. & Goldberg, A.L. (1985). Production of abnormal proteins in *E.coli* stimulates transcription of *lon* and other heat shock genes. *Cell* 41, 587-595.
- Goloubinoff, P., Gatenby, A.A. & Lorimer, G.H. (1989). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose biphosphate carboxylase oligomers in *Escherichia coli*. *Nature* 337, 44-47.
- Grantham, R. (1980). Workings of the genetic code. *Trends Biochem. Sci.* 5, 327-331.
- Grantham, R., Gautier, C. & Gouy, M. (1980). Codon frequencies of 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucl. Acids Res.* 9, 243-274.
- Grantham, R., Perrin, P. & Mouchiroud, D. (1986). Patterns in codon usage of different kinds of species. In *Oxford Survey in Evolutionary Biology* (Eds. Dawkins, R. & Ridley, M.) Oxford University Press vol. 3, pp. 48-81.
- Grossman, A., Bartlett, S. & Chua, N-H. (1980). Energy-dependent uptake of cytoplasmically synthesized polypeptides by chloroplasts. *Nature* 285, 625-628.
- Gurevitz, M., Sommerville, C.R. & McIntosh, L. (1985). Pathway of assembly of ribulose bisphosphate carboxylase/oxygenase from *Anabaena 7120* expressed in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 82, 6546-6550.
- Gutteridge, S. & Gatenby, A.A. (1987). The molecular analysis of the assembly, structure and function of Rubisco. In *Oxford Surveys of Plant Molecular and Cell Biology*, Vol 4, ed. B.J. Miflin) Oxford University Press pp 95-135.
- Gutteridge, S., Phillips, A.L., Kettleborough, C.A., Parry, M.A.J. & Keys, A.J. (1986). Expression of bacterial Rubisco genes in *Escherichia coli*. *Phil. Trans. R. Soc. Lond. B* 313, 433-445.
- Haas, I.G. & Wabl, M. (1983). Immunoglobulin heavy chain binding protein. *Nature* 306, 387-389.
- Hampp, R., Goller, M., Ziegler, H. (1982). Adenylate levels, energy charge, and phosphorylation potential during dark-light and light-dark transition in chloroplast, mitochondria, and cytosol of mesophyll protoplasts from *Avena sativa* L. *Plant Physiol.* 69, 448-455.



- Hartman, F.C., Stringer, C.D., Omnaas, J., Donnelly, M.I. & Fray, B. (1982). Purification and sequencing of cyanogen bromide fragments from ribulose biphosphate carboxylase/oxygenase from *Rhodospirillum rubrum*. Arch. Biochem. Biophys. 219, 422-437.
- Hedrick, J.L. & Smith, A.J. (1968). Size and charge isomer separation and estimation of molecular weights of proteins by disc gel electrophoresis. Arch. Biochem. Biophys. 126, 155-164.
- Heinhorst, S. & Shively, J.M. (1983). Encoding of both subunit of ribulose-1.5-bisphosphate carboxylase by organelle genome of *Cyanophora paradoxa*. Nature 304, 373-374.
- Hemmingsen, S.M. & Ellis, R.J. (1986). Purification and properties of ribulose biphosphate carboxylase large subunit binding protein. Plant Physiol. 80, 269-276.
- Hemmingsen, S.M., Woolford, C., van der Vies, S.M., Tilly, K., Dennis, D.T., Georgopoulos, C.P., Hendrix, R.W. & Ellis, R.J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature 333, 330-334.
- Hendrix, R.W. & Tsui, L. (1978). Role of the host in virus assembly: cloning of the *Escherichia coli* *groE* gene and identification of its protein product. Proc. Natl. Acad. Sci. USA 75, 136-139.
- Hendrix, R.W. (1979). Purification and properties of *groE*, a host protein involved in bacteriophage assembly. J. Mol. Biol. 129, 375-392.
- Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28, 351-359.
- Herrera-Estrella, L., van den Broeck, G., Maenhaut, R., van Montagu, M., Schell, J., Timko, M. & Cashmore, A.R. (1984). Light inducible and chloroplast associated expression of a chimaeric gene introduced into *Nicotiana tabacum* using a Ti plasmid vector. Nature 310, 115-120.
- Highfield, P.E. & Ellis, R.J. (1978). Synthesis and transport of the small subunit of chloroplast ribulose biphosphate carboxylase. Nature 271, 420-424.
- Hindersson, P., Petersen, C.S., Pedersen, N.S., Høiby, N. & Axelsen, N.H. (1984). Immunological cross-reaction between antigen Tp-4 of *Treponema pallidum* and an antigen common to a wide range of bacteria. Acta Path. Immunol. Scand. sect. B 92, 183-188.
- Hirai, A. & Wildman, S.G. (1977). Kinetic analysis of fraction 1 protein biosynthesis in young protoplast of tobacco leaves. Biochim. Biophys. Acta 439, 39-52.

- Hjelmeland, L.M. (1980). A nondenaturing zwitterionic detergent for membrane biochemistry: design and synthesis. *Proc. Natl. Acad. Sci. USA* 77, 6368-6370.
- Hoekema, A., Kastelein, R.A., Vasser, M. & de Boer, H.A. (1987). Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: Experimental approach to study the role of biased codon usage in gene expression. *Mol. Cell. Biol.*, 7, 2914-2924.
- Hohn, B. & Hohn, T. (1974). Activity of empty, headlike particles for packaging of DNA of bacteriophage  $\lambda$  *in vitro*. *Proc. Natl. Acad. Sci. USA* 71, 2372-2376.
- Hohn, B., Wurtz, M., Klein, B., Lustig, A. & Hohn, T. (1974). Phage lambda DNA packaging *in vitro*. *J. Supramol. Struct.* 2, 302-317.
- Høiby, N. (1975). Cross-reactions between *Pseudomonas aeruginosa* and thirty-six other bacterial species. *Scand. J. Immunol.* 4, Suppl. 2, 187-196.
- Holmes, D.S. & Quigley, M. (1981). A rapid method for the preparation of bacterial plasmids. *Anal. Biochem.* 114, 193-197.
- Hopp, T.P. & Woods, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* 78, 3824-3828.
- Howe, C.J., Bowman, C.M., Dyer, T.A. & Gray, J.C. (1982). Localization of wheat chloroplast genes for the beta and epsilon subunits of ATP synthase. *Mol. Gen. Genet.* 186, 525-530.
- Hubbard, S.C. & Ivatt, R.J. (1981). Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 50, 555-583.
- Hunt, A.G., Chu, N.M., Odell, J.T., Nagy, F. & Chua, N-H. (1987). Plant cells do not properly recognize animal gene polyadenylation signals. *Plant Mol. Biol.* 8, 23-25.
- Hurt, E.C., Goldschmidt-Clermont, M., Pesold-Hurt, B., Rochaix, J-D. & Schatz, G. (1986). A mitochondrial presequence can transport a chloroplast-encoded protein into yeast mitochondria. *J. Biol. Chem.* 261, 11440-11443.
- Huynh, T.V., Young, R.A. & Davis, R.W. (1985). Constructing and screening cDNA libraries in  $\lambda$ gt10 and  $\lambda$ gt11. In: *DNA cloning: A practical approach*, ed. D.M. Glover, IRL Press, Oxford, pp. 49-78.
- Hwang, D.S. & Kaguni, J.M. (1988a). Purification and characterization of the *dnaA46* gene product. *J. Biol. Chem.* 263, 10625-10632.

- Hwang, D.S. & Kaguni, J.M. (1988b). Interaction of *dna46* protein with a stimulatory protein in replication from the *Escherichia coli* chromosomal origin. *J. Biol. Chem.* 263, 10633-10640.
- Incharoensakdi, A. Takabe, T., Takabe, T. & Akazawa, T. (1985). Heterologous hybridization of ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) restores the enzyme activities. *Biochem. Biophys. Res. Commun.* 126, 698-704
- IUPAC-IUB Commission on Biochemical Nomenclature (CBN) (1970). Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. *Biochem. J.* 120, 449-454.
- IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBM) (1984). Nomenclature and symbolism for amino acids and peptides. *Biochem. J.* 219, 345-373.
- Jameson, B.A. & Wolf, H. (1988). The antigenic index: a novel algorithm for predicting antigenic determinants CABIOS 4, 181-186.
- Jenkins, A.J., March, J.B., Oliver, I.R. & Masters, M. (1986). A DNA fragment containing the *groE* genes can suppress mutations in the *Escherichia coli dnaA* gene. *Mol. Gen. Genet.* 202, 446-454.
- Jindal, S., Dudani, A.K., Singh, B., Harley, C.B. & Gupta, R.S. (1989). Primary structure of a human mitochondrial protein homologous to the bacterial and plant chaperonins and the 65-kilodalton mycobacterial antigen. *Mol. Cell. Biol.* 9, 2279-2283.
- Johnson, R.A. (1987). Properties and synthesis of the ribulose 1,5 bisphosphate carboxylase large subunit binding protein of *hordeum vulgare*. Ph.D. Thesis, University of Warwick, U.K.
- Johnston, A. & Thrope, R. (1982). *Immunochemistry in practice*, Blackwell Scientific publications, Oxford, pp. 27-40.
- Jordan, D.B. & Ogren, W.L. (1981). Species variation in kinetic properties of ribulose-1,5-bisphosphate carboxylase-oxygenase. *Nature* 291, 513-515.
- Jordan, D.B. & Ogren, W.L. (1983). Specific variation in kinetic properties of ribulose-1,5-bisphosphate carboxylase-oxygenase. *Arch. Biochim. Biophys.* 227, 425-433.
- Joshi, C.P. (1987). Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. *Nucl. Acids Res.* 15, 9627-9640.
- Kaijser, B. (1975). Immunological studies of an antigen common to many gram-negative bacteria with special reference to *E.coli*. *Intern. Arch. Allergy Appl. Immun.* 48, 72-81.

- Kaiser, A.D., Syvanen, M. & Masuda, T. (1974). Processing and assembly of the head of bacteriophage lambda. *J. Supramol. Struct.* 2, 318-328.
- Kaiser, A.D., Syvanen, M., & Masuda, T. (1975). DNA packaging steps in bacteriophage lambda head assembly. *J. Mol. Biol.* 91., 175-186.
- Kakidani, H., Furutani, Y., Takahashi, H., Noda, M., Morimoto, Y., Hirose, T., Asai, M., Iyanayama, S., Nakanishi, S., and Numa, S. (1982). Cloning and sequence analysis of cDNA for porcine b-neo-endorphin /dynorphin precursor. *Nature* 298, 245-249.
- Kawashima, N. & Wildman, S.G. (1972). Studies on fraction I protein. IV Mode of inheritance of primary structure in relation to whether chloroplast or nuclear DNA contains the code a chloroplast protein. *Biochim. Biophys. Acta* 262, 42-49.
- Keegstra, K. (1989). Transport and routing of proteins into chloroplasts. *Cell* 56, 247-253.
- Keith, B. & Chua, N.-H. (1986). Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco. *EMBO J.* 5, 2419-2425.
- Kimball, B.A. (1983). Carbon dioxide and agricultural yield: An assemblage and, analysis of 430 prior observations. *Agronomy J.* 75, 779-788.
- Kirk, J.T.O. & Tilney-Bassett, R.A.E. (1978). The Plastids: Their chemistry, structure, growth and inheritance. Amsterdam/New York/Oxford: Elsevier/North-Holland 960 pp. 2nd ed.
- Klein, R.R. & Mullet, J.E. (1987). Control of gene expression during higher plant chloroplast development. *J Biol. Chem.* 262, 4341-4348.
- Kochan, J. & Murialdo, H. (1983). Early intermediates in bacteriophage lambda prohead assembly. II. Identification of biological active intermediates. *Virology* 131, 100-115.
- Kochan, J., Carrascosa, J.L. & Murialdo, H. (1984). Bacteriophage lambda preconnectors; purification and structure. *J. Mol. Biol.* 174, 433-447.
- Kohli, J. & Grosjean, H. (1981). Usage of the three termination codons: compilation and analysis of the known eukaryotic and prokaryotic translation termination sequences. *Mol. Gen. Genet.* 182, 430-439.
- Koller, B. Gingrich, J.C., Steigler, G.L., Farley, M.A., Delius, H. & Hallick, R.B. (1984). Nine introns with conserved boundary sequences in the *Euglena gracilis* chloroplast ribulose-1.5-bisphosphate carboxylase gene. *Cell* 36, 545-553.

- Kozak, M. (1983). Comparison of initiation of protein: synthesis in prokaryotes, eukaryotes and organelles. *Microbiol. Rev.* 47, 1-45.
- Kozak, M. (1984). Point mutations close to the AUG initiation codon affect the efficiency of translation of rat preproinsulin *in vivo*. *Nature* 308, 241-246.
- Krieg, P.A. & Melton, D.A. (1987). In vitro RNA synthesis with SP6 RNA polymerase. *Meth. in Enzymol.*, 155, 397-415.
- Kuwano, M., Ono, M., Endo, H., Hori, K., Nakamura, K., Hirota, Y., & Ohnishi, Y. (1977). Gene affecting longevity of messenger RNA: a mutant of *Escherichia coli* with altered mRNA stability. *Mol. Gen. Genet.* 154, 279-285.
- Kyte, J. & Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105-132.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Laing, W.A., Ogren, W.L. & Hageman, R.H. (1974). Regulation of soybean net photosynthetic CO<sub>2</sub> fixation by the interaction of CO<sub>2</sub>, O<sub>2</sub> and ribulose-1,5-diphosphate carboxylase. *Plant Physiol.* 54, 678-685.
- Lamb, F.I. (1984). Castor bean lectins: construction and sequencing of cDNA clones. Ph.D. Thesis, University of Warwick, England.
- Lamb, J.R. & Young, D.B. (1989). T-cell recognition of stress proteins: a link between infectious and autoimmune disease. In *Molecular Biology and Medicine*, an issue on autoimmune diseases (in press).
- Lamppa, G.K. & Bendich, A.J. (1979). Changes in chloroplast DNA levels during development of pea (*Pisum sativum*). *Plant Physiol.* 64, 126-130.
- Landridge, P. (1981). Synthesis of the large subunit of spinach ribulose biphosphate carboxylase may involve a precursor polypeptide. *FEBS lett.* 123, 85-89.
- Landry, S.J. & Bartlett, S.G. (1989). The small subunit of ribulose-1,5-biphosphate carboxylase/oxygenase and its precursor expressed in *Escherichia coli* are associated with the *groEL* protein. *J. Biol. Chem.* 264, 9090-9093.

- Lang, K., Schmidt, F.X. & Fischer, G. (1987). Catalysis of protein folding by propylisomerase. *Nature* 329, 268-270.
- Larimer, F.W., Machanoff, R. & Hartman, F.M. (1986). A reconstitution of the gene for ribulose biphosphate carboxylase from *Rhodospirillum rubrum* that expressed the authentic enzyme in *Escherichia coli*. *Gene* 41, 113-120.
- Laskey, R.A. & Earnshaw, W.C. (1980). Nucleosome assembly (review article). *Nature* 286, 763-767.
- Laskey, R.A., Honda, B.M., Mills, A.D. & Finch, J.T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfer s them to DNA. *Nature* 275, 416-420.
- Leaver, C.J., Hack, E. & Forde, B.G. (1983). Protein synthesis by isolated plant mitochondria. *Meth. in Enzymol.* 97, 476-484.
- Lecker, S., Lill, R., Ziegelkoffer, T., Bassford, Jr., P.J., Kumamoto, C.A. & Wickner, W. (1989). Three pure chaperone proteins of *Escherichia coli*, sec B, trigger factor and *groEL*, from soluble complexes with precursor proteins *in vitro*. *EMBO J.* 8, 2703-2709.
- Lehninger, A.L. (1970). *Biochemistry*, Worth Publishers Inc.
- Lehrah, H., Diamond, D., Wozney, J.M. & Boedtker, H. (1977). RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* 16, 4743-4751.
- Lennox, C.R. & Ellis, R.J. (1986). The carboxylase large subunit binding protein: photoregulation and reversible dissociation. *Biochem. Soc. Trans.* 14, 9-11.
- Lindquist, S. & Graig, E.A. (1988). The heat-shock proteins. *Ann. Rev. Genet.* 22, 631-677.
- Link, G., Coen, D.M. & Bogorad, L. (1978). Differential expression of the gene for the large subunit of ribulose biphosphate carboxylase in maize leaf cell types. *Cell* 15, 725-731.
- Lipman, D.J. & Pearson, W.R. (1985). Rapid and sensitive protein similarity searches. *Science* 227, 1435-1441.
- Liren, L., Cuifang, L. & Renbang, X. (1988). Dissociation and reconstitution of immobilized ribulose-1.5-biphosphate carboxylase from tobacco. *Scientia Sinica (Series B)* 31, 1204-1212.
- Lorimer, G.H. (1981). Ribulose biphosphate carboxylase-amino acid sequence of a peptide bearing the activator carbon dioxide. *Biochemistry, Wash.* 20, 1236-1240.

- Lorimer, G.H. & Miziorko, H.M. (1980). Carbamate formation on the  $\epsilon$ -amino group of a lysyl residue as the basis for the activation of ribulose-bisphosphate carboxylase by  $\text{CO}_2$  and  $\text{Mg}^{2+}$ . *Biochemistry, Wash.* 19, 5321-5328.
- Lorimer, G.H., Badger, M.R. & Andrews, T.J. (1976). The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism and physiological implications. *Biochemistry, Wash.* 15, 529-536.
- Lorimer, G.H., Badger, M.R. & Andrews, T.J. (1977). D-Ribulose-1,5-bisphosphate carboxylase-oxygenase: Improved methods of the activation and assay of catalytic activities. *Anal. Biochem.* 78, 66-75.
- Lorimer, G.H., Andrews, T.J., Pierce, J. & Schloss, J.V. (1986). 2'-carboxy-3-keto-D-arabinitol 1,5-bisphosphate, the six-carbon intermediate of the ribulose bisphosphate carboxylase reaction. *Phil. Trans. R. Soc. Lond. B* 313, 397-407.
- Lubben, T.H., Donaldson, G.K., Viitanen, P.V. & Gatenby, A.A. (1989). Several proteins imported into chloroplasts form stable complexes with the groEL-related chloroplast molecular chaperone. *Plant Cell* (in press).
- Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. & Scheele, G.A. (1987). Selection of AUG initiation codons differs in plants and animals. *EMBO J.* 6, 43-48.
- Lycett, G.W., Delaney, A.J. & Croy, R.D. (1983). Are plant genes different? *FEBS Lett.* 153, 43-46.
- MacLennan, D.H., Brandl, C.J., Korczak, B. & Green, N.M. (1985). Amino-acid sequence of  $\text{Ca}^{2+}$  +  $\text{Mg}^{2+}$ -dependent ATPase from rabbit muscle sarcoplasmic reticulum, deduced from its complementary DNA sequence. *Nature* 316, 696-700.
- Maizel, Jr., J.V. & Lenk, R.P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA* 78, 7665-7669.
- Mandel, M. and Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* 53, 159-162.
- Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982). *Molecular cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York.
- Markwell, M.A.K. (1982). A new solid-state reagent for iodination of proteins. 1. Conditions for the efficient labeling of antiserum. *Anal. Biochem.* 127, 427-432.
- Marmur, J. and Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.* 5, 109-118.



- Maruyama, T., Gojobori, T., Aota, S., & Ikemura, T. (1986). Codon usage tabulated from the GenBank genetic sequence data. *Nucl. Acids Res.* *14*, r151-r189.
- Matsuoka, M., Kano-Murakami, Y., Tanaka, Y., Ozeki, Y. & Yamamoto, N. (1987). Nucleotide sequence of cDNA encoding the small subunit of ribulose-1,5-bisphosphate carboxylase from maize. *J. Biochem.* *102*, 673-676.
- Mazur, B.J. & Chui, C-F. (1985). Sequence of a genomic DNA clone for the small subunit of ribulose-1.5-bisphosphate carboxylase-oxygenase from tobacco. *Nucl. Acids Res.* *13*, 2373-2386.
- McFadden, B.A., Torres-Ruiz, J.A. & Franceschi, V.R. (1989). Localization of ribulose-biphosphate carboxylase-oxygenase and it putative binding protein in the cell envelope of *chromatium vinosum*. *Planta* *178*, 297-302.
- McGrath, J.P., Capon, D.J., Goeddel, D.V. & Levinson, A.D. (1984). Comparative biochemical properties of normal and activated hum ras p21 protein. *Nature*, *310*, 644-649.
- McIntosh, L. Poulsen, C. & Bogorad, L. (1980). Chloroplast gene sequence for the large subunit of ribulose bisphosphate carboxylase of maize. *Nature* *288*, 556-560.
- McMullin, T.W. & Hallberg, R.L. (1987). A normal mitochondrial protein is selectively synthesized and accumulated during heat shock in *Tetrahymena thermophila*. *Mol. Cell. Biol.* *7*, 4414-4423.
- McMullin, T.W. & Hallberg, R.L. (1988). A highly evolutionarily conserved mitochondrial protein is structurally related to the protein encoded by the *Escherichia coli groEL* gene. *Mol. Cell. Biol.* *8*, 371-380.
- Mehra, V., Sweetser, D., & Young, R.A. (1986). Efficient mapping of protein antigenic determinants. *Proc. Natl. Acad. Sci. USA* *83*, 7013-7017.
- Meinkoth, J. and Wahl, G. (1984). Hybridization of nucleic acids immobilized on solid supports. *Anal. Biochem.* *138*, 267-284.
- Miki, T., Orita, T., Furuno, M. & Horiuchi. (1988). Control of cell division by sex factor F in *Escherichia coli* III. Participation of the *groES* (mopB) gene of the host bacteria. *J. Mol. Biol.* *201*, 327-338.
- Miller, J.H. (1972). Experiment 57: Purification of b-galactosidase. In: *Experiments in molecular genetics*, Cold Spring Harbor Laboratory, New York, pp. 403-404.

- Milos, P. & Roy, H. (1984). ATP-released large subunits participate in the assembly of RuBP carboxylase. *J. Cell. Biochem.* 24, 153-162.
- Milos, P., Bloom, M.V. & Roy, H. (1985). Methods for studying the assembly of ribulose biphosphate carboxylase. *Plant Mol. Reporter* 3, 33-42.
- Mishkind, M.L., Wessler, S.R. & Schmidt, G.W. (1984). Functional determinants in transit sequences: import and partial maturation by vascular plant chloroplasts of the ribulose-1,5-bisphosphate carboxylase small subunit of *Chlamydomonas*. *J. Cell Biol.* 100, 226-234.
- Miziorko, H.M. & Lorimer, G.H. (1983). Ribulose-1.5-bisphosphate carboxylase-oxygenase. *Ann. Rev. Biochem.* 52, 507-535.
- Molineux, I.J. & Gefner, M.L. (1979). Properties of the *Escherichia coli* DNA binding (unwinding) protein: interaction with DNA polymerase and DNA. *Proc. Natl. Acad. Sci. USA* 71, 3858-3862.
- Molineux, I.J. & Gefner, M.L. (1975). Properties of the *Escherichia coli* DNA-binding (unwinding) protein interaction with nucleolytic enzymes and DNA. *J. Mol. Biol.* 98, 811-825.
- Morgenthaler, J.J., Marsden, M.P.F. & Price, C.A. (1975). Factors affecting the separation of photosynthetically competent chloroplasts in gradients of silica sols. *Arch. Biochem. Biophys.* 168, 289-301.
- Munro, S. & Pelham, H.R.B. (1986). An Hsp70-like protein in the ER: identity with the 78 kd glucose-regulated protein and the immunoglobulin heavy chain binding protein. *Cell* 46, 291-300.
- Murialdo, H. (1979). Early intermediates in bacteriophage lambda prohead assembly. *Virology* 96, 341-367.
- Murialdo, H. & Becker, A. (1978). A genetic analysis of bacteriophage lambda prohead assembly *in vitro*. *J. Mol. Biol.* 125, 57-74.
- Murialdo, H. & Siminovitch, L. (1972a). The morphogenesis of bacteriophage lambda. IV. Identification of gene products and control of the expression of the morphogenetic information. *Virology* 48, 785-823.
- Murialdo, H. & Siminovitch, L. (1972b). The morphogenesis of bacteriophage lambda. V. Form-determining function of the genes required for assembly of the head. *Virology* 48, 824-835.
- Murphy, G. & Kavanagh, T. (1988). Speeding-up the sequencing of double-stranded DNA. *Nucl. Acids Res.* 16, 5198.
- Murray, E.E., Lotzer, J. & Eberle, M. (1989). Codon usage in plant genes. *Nucl. Acids Res.* 17, 477-498.

- Milos, P. & Roy, H. (1984). ATP-released large subunits participate in the assembly of RuBP carboxylase. *J. Cell. Biochem.* 24, 153-162.
- Milos, P., Bloom, M.V. & Roy, H. (1985). Methods for studying the assembly of ribulose biphosphate carboxylase. *Plant Mol. Reporter* 3, 33-42.
- Mishkind, M.L., Wessler, S.R. & Schmidt, G.W. (1984). Functional determinants in transit sequences: import and partial maturation by vascular plant chloroplasts of the ribulose-1,5-bisphosphate carboxylase small subunit of *Chlamydomonas*. *J. Cell Biol.* 100, 226-234.
- Miziorko, H.M. & Lorimer, G.H. (1983). Ribulose-1.5-bisphosphate carboxylase-oxygenase. *Ann. Rev. Biochem.* 52, 507-535.
- Molineux, I.J. & Gefner, M.L. (1979). Properties of the *Escherichia coli* DNA binding (unwinding) protein: interaction with DNA polymerase and DNA. *Proc. Natl. Acad. Sci. USA* 71, 3858-3862.
- Molineux, I.J. & Gefner, M.L. (1975). Properties of the *Escherichia coli* DNA-binding (unwinding) protein interaction with nucleolytic enzymes and DNA. *J. Mol. Biol.* 98, 811-825.
- Morgenthaler, J.J., Marsden, M.P.F. & Price, C.A. (1975). Factors affecting the separation of photosynthetically competent chloroplasts in gradients of silica sols. *Arch. Biochem. Biophys.* 168, 289-301.
- Munro, S. & Pelham, H.R.B. (1986). An Hsp70-like protein in the ER: identity with the 78 kd glucose-regulated protein and the immunoglobulin heavy chain binding protein. *Cell* 46, 291-300.
- Murialdo, H. (1979). Early intermediates in bacteriophage lambda prohead assembly. *Virology* 96, 341-367.
- Murialdo, H. & Becker, A. (1978). A genetic analysis of bacteriophage lambda prohead assembly *in vitro*. *J. Mol. Biol.* 125, 57-74.
- Murialdo, H. & Siminovitch, L. (1972a). The morphogenesis of bacteriophage lambda. IV. Identification of gene products and control of the expression of the morphogenetic information. *Virology* 48, 785-823.
- Murialdo, H. & Siminovitch, L. (1972b). The morphogenesis of bacteriophage lambda. V. Form-determining function of the genes required for assembly of the head. *Virology* 48, 824-835.
- Murphy, G. & Kavanagh, T. (1988). Speeding-up the sequencing of double-stranded DNA. *Nucl. Acids Res.* 16, 5198.
- Murray, E.E., Lotzer, J. & Eberle, M. (1989). Codon usage in plant genes. *Nucl. Acids Res.* 17, 477-498.

- Ogren, W.L. (1984). Photorespiration: pathways, regulation and modification. *Ann. Rev. Plant. Physiol.* 35, 415-442.
- Opperham, T., Hong, T.-H. & Surzycki, S.J. (1989). Chloroplast and nuclear genomes of *Chlamydomonas reinhardtii* share homology with *Escherichia coli* genes for DNA replication, and transition. *Curr. Genet.* 15, 39-46.
- Pain, R.H. (1987). A case of mistaken identity. *Nature* 328, 298.
- Pain, D. & Blobel, G. (1987). Protein import into chloroplasts requires a chloroplast ATPase. *Proc. Natl. Acad. Sci. USA* 84, 3288-3292.
- Pain, D., Kanwar, Y.S. & Blobel, G. (1988) Identification of a receptor for protein import into chloroplasts and its localization to envelope contact zones. *Nature* 331, 232-237.
- Parkinson, J.S. (1968). Genetics of the left arm of the chromosome of bacteriophage lambda. *Genetics.* 59, 311-325.
- Pelham, H.R.B. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell* 46, 959-961.
- Pelham, H.R.B., (1988). Coming in from the cold. *Nature* 332, 776-777.
- Perchorowicz, J.T. & Jensen, R.G. (1983). Photosynthesis and activation of ribulose bisphosphate carboxylase in wheat seedlings. *Plant Physiol.* 71, 955-960.
- Pfisterer, J. Lachmann, P. & Klopstech, K. (1982). Transport of proteins into chloroplasts. Binding of nuclear-encoded chloroplast proteins to the chloroplast envelope. *Eur. J. Biochem.* 126, 143-148.
- Pon, N.G., Rabin, B.R. & Calvin, M. (1963). Mechanism of the carboxydismutase reaction. I. The effect of premilary incubation of substrates, metal ion and enzyme on activity. *Biochem. Z.* 338, 7-19.
- Poulsen, C. (1981). Comments on the structure and function of the large subunit of the enzyme ribulose bisphosphate carboxylase-oxygenase. *Carlsberg Res. Commun.* 46, 259-278.
- Prasad, T.K. & Hallberg, R.L. (1989). Identification and metabolic characterization of the *Zea mays* mitochondrial homolog of the *Escherichia coli* *groEL* protein. *Plant Mol. Biol.* 12, 609-618.
- Proudfoot, N.J. & Brownlee, G.G. (1976) 3' non-coding region sequences in eukaryotic messenger RNA, *Nature* 263, 211-214.

- Pushkin, A.V., Tsuprun, V.L., Solovjeva, N.A., Shubin, V.V., Evstigneeva, Z.G. & Kretovich, W.L. (1982). High molecular weight pea leaf protein similar to the *groE* protein of *Escherichia coli*. *Biochim. Biophys. Acta* 704, 379-384.
- Quigley, G.J., Gehrke, L., Roth, D.A. & Auron, P.E. (1984). Computer-aided nucleic acid secondary modeling incorporating enzymatic digestion data. *Nucl. Acids Res.* 12, 347-366.
- Quivey, R.G. & Tabita, F.R. (1984). Cloning and expression in *Escherichia coli* of the form II ribulose 1.5-bisphosphate carboxylase/oxygenase gene from *Rhodospseudomonas sphaeroides*. *Gene* 31, 91-101.
- Ray, P.N. & Murialdo, H. (1975). The role of gene Nu3 in bacteriophage lambda head morphogenesis. *Virology* 64, 247-263.
- Reading, D.S., Hallberg, R.L. & Myers, A.M. (1989). Characterization of the yeast HSP60 gene coding for a mitochondrial assembly factor. *Nature* 337, 655-659.
- Reith, M.E. & Cattolico, R.A. (1986). Inverted repeat of *Olishodiscus luteus* chloroplast DNA contains genes for both the subunits of ribulose-1.5-bisphosphate carboxylase and the 32000-dalton QB protein: phylogenetic implications. *Proc. Natl. Acad. Sci. USA* 83, 8599-8603.
- Renart, J., Reiser, J. & Stark, G.R. (1979). Transfer of proteins from gels to diazobenzyl-oxymethyl-paper and detection with antisera: A method for studying antibody specificity and antigen structure. *Proc. Natl. Acad. Sci. USA* 76, 3116-3120.
- Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* 113, 237-251.
- Robinson, C. & Ellis, R.J. (1984). Transport of proteins into chloroplasts. The precursor of small subunit of ribulose bisphosphate carboxylase is processed to the mature size in two steps. *Eur. J. Biochem.* 142, 343-346.
- Robinson, C. & Ellis, R.J. (1985). Transport of proteins into chloroplasts. The effect of incorporation of amino acid analogues on the import and processing of chloroplast polypeptides. *Eur. J. Biochem.* 152, 67-73.
- Rochaix, J-D. (1985). Genetic organisation of the chloroplast. *Int. Rev. Cytol.* 93, 57-91.
- Rothman, J.E. & Schmid, S.L. (1986). Enzymatic recycling of clathrin from coated vesicles. *Cell* 46, 5-9.

- Roy, H., Adari, H. & Costa, K.A. (1979). Characterization of the free subunits of ribulose-1.5-bisphosphate carboxylase. *Plant Sci. Lett.* 16, 305-318.
- Roy, H., Bloom, M., Milos, P. & Monroe, M. (1982). Studies on the assembly of the large subunit of ribulose biphosphate carboxylase in isolated pea chloroplasts. *J. Cell Biol.* 94, 20-27.
- Roy, H., Cannon, S. & Gilson, M. (1988). Review: Assembly of Rubisco native subunits. *Biochim. Biophys. Acta* 957, 323-334.
- Roy, H., Costa, K.A. & Adari, H. (1978). Free subunits of ribulose-1.5-bisphosphate carboxylase in pea leaves. *Plant Sci. Lett.* 11, 159-168.
- Roy, H., Hubbs, A. & Cannon, S. (1988). Stability and dissociation of the large subunit Rubisco binding protein complex *in vitro* and *in organello*. *Plant Physiol.* 86, 50-53.
- Ruben, S.M., VanDenBrink-Webb, S.E., Rein, D.C. & Meijer, R.R. (1988). Suppression of the *Escherichia coli* ssb-1 mutation by an allele of *groEL*. *Proc. Natl. Acad. Sci. USA* 85, 3767-3771.
- Rusche, J.R. & Howard-Flanders, P. (1985). Hexamine cobalt chloride promotes intermolecular ligation of blunt end DNA fragments by T4 DNA ligase. *Nucl. Acids Res.* 13, 1997-2008.
- Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
- Sakano, K., Kung, S.D. & Wildman, S.G. (1974). Identification of several chloroplast DNA genes which code for the large subunit of *Nicotiana* fraction 1 protein. *Mol. Gen. Genet.* 130, 91-97.
- Salvucci, M.E., Portis, A.R.Jr & Ogren, W.L. (1985). A soluble chloroplast protein catalyzes ribulose biphosphate carboxylase/oxygenase activation *in vitro*. *Photosynth. Res.* 7, 193-201.
- Salvucci, M.E., Portis, A.R.Jr & Ogren, W.L. (1986). Light and CO<sub>2</sub> response of ribulose-1.5-bisphosphate carboxylase/oxygenase activation in *Arabidopsis* leaves. *Plant Physiol.* 80, 655-659.
- Savakis, C., Demetri, G. & Cherbas, P. (1980). Ecdysteroid-inducible polypeptides in a *Drosophila* cell line. *Cell* 22, 665-674.
- Schmidt, G.W. & Mishkind, M.L. (1983). Rapid degradation of unassembled ribulose-1,5-bisphosphate carboxylase small subunit from *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA* 80, 2632-2636.

- Schmidt, M.G., Rollo, E.E., Grodberg, J. & Oliver, D.B. (1988). Nucleotide sequence of the *secA* gene and *secA* (Ts) mutations preventing protein export in *Escherichia coli*. *J. Bact.* 170, 34-4-3414.
- Schneider, G. Linqvist, Y., Brändén, C-I & Lorimer, G.H. (1986). The three-dimensional structure of ribulose-1.5-bisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum* at 2.9Å resolution. *EMBO J.* 5, 3409-3415.
- Sekimizu, K., Bramhill, D. & Kornberg, A. (1987). ATP activates *dnaA* protein in initiating replication of plasmids bearing the origin of the *E.coli* chromosome. *Cell* 50, 259-265.
- Sellers, P.H. (1974). On the theory and computation of evolutionary distances. *SIAM J. Appl. Math.* 26, 787-793.
- Sharma, S., Rodgers, L. & Brandsma, J. (1985). SV40 T antigen and the exocytotic pathway. *EMBO J.* 4, 1479-1489.
- Shinnick, T.M. (1987). The 65-kilodalton antigen of *Mycobacterium tuberculosis*. *J. Bact.* 169, 1080-1088.
- Shinnick, T.M., Plikaytis, B.B., Hyshe, A.D., van Landingham, R.M. & Walker, L.L. (1989). The *Mycobacterium tuberculosis* BCG-a protein has homology with the *Escherichia coli* *groES* protein. *Nucl. Acids Res.* 17, 1254.
- Shinozaki, K. & Sugiura, M. (1982). The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1.5-bisphosphate carboxylase/oxygenase. *Gene* 20, 91-102.
- Shinozaki, K. & Sugiura, M. (1983). The gene for the small subunit of ribulose-1.5-bisphosphate carboxylase/oxygenase is located close to the gene for the large subunit in the cyanobacterium *Anacystis nidulans* 6301. *Nucl. Acids Res.* 11, 6957-6964.
- Shinozaki, K. & Sugiura, M. (1985). Genes for the large and small subunit of ribulose bisphosphate carboxylase/oxygenase constitute a single operon in a cyanobacterium *Anacystis nidulans* 6301. *Mol. Gen. Genet.* 200, 27-32.
- Simon, L.D., Tomczak, K. & St. John, A.C. (1978). Bacteriophage inhibits degradation of abnormal proteins in *E.coli*. *Nature* 275, 424-428.
- Simon, L.D., Randolph, B., Irwin, N. & Binkowski, G. (1983). Stabilization of proteins by a bacteriophage T4 gene cloned in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 80, 2059-2062.
- Smith, S.M. & Ellis, R.J. (1979). Processing of the small subunit precursor of ribulose bisphosphate carboxylase and its assembly into the whole enzyme are stromal events. *Nature* 278, 662-664.



- Smith, S.M. & Ellis, R.J. (1981). Light-stimulated accumulation of transcripts of nuclear and chloroplast genes for ribulose biphosphate carboxylase. *J. Mol. Appl. Genet.* 1, 127-137.
- Somerville, C.R. & Ogren, W.L. (1982). Genetic manipulation of photorespiration. *Trends Biochem. Sci.* 7, 171-174.
- Somerville, C.R. & Somerville, S.C. (1984). Cloning and expression of *Rhodospirillum rubrum* ribulose 1.5-bisphosphate carboxylase gene in *E.coli*. *Mol. Gen. Genet.* 193, 214-219.
- Somerville, C.R., Portis, A.R. Jr & Ogren, W.L. (1982). A mutant of *Arabidopsis thaliana* which lacks activation of RuBP carboxylase *in vivo*. *Plant Physiol.* 70, 381-387.
- Somerville, C., Fitchen, J., Somerville, S., McIntosh, L. & Nargang, F. (1983). Enhancement of net photosynthesis by genetic manipulation of photorespiration and RuBP carboxylase/oxygenase. In: *Advances in Gene Technology: Molecular Genetics of Plants and Animals* (eds. K. Dourney, R.W. Voellmy, F. Ahmed, J. Schultz) New York pp 295-309.
- Somerville, C.R., McIntosh, L., Fitchen, J. & Gurevitz, M. (1986). The cloning and expression in *Escherichia coli* of RuBP carboxylase/oxygenase large subunit genes. *Meth. Enzymol.* 118, 419-433.
- Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503-517.
- Spalding, M.H., Spreitzer, R.J. & Ogren, W.L. (1983). Reduced inorganic carbon transport in a CO<sub>2</sub>-requiring mutant of *Chlamydomonas reinhardtii*. *Plant Physiol.* 73, 273-276.
- Spreitzer, R.J. & Mets, L.J. (1980). Non-mendelian mutation affecting ribulose-1.5-bisphosphate carboxylase structure and activity. *Nature* 285, 114-115.
- Stancel, G.M. & Deal, Jr. W.C., (1968). Metabolic control and structure of glycolytic enzymes. Dissociation of yeast glyceraldehyde-3-phosphate dehydrogenase into subunits by ATP. *Biochem. Biophys. Res. Commun.* 31, 398-403.
- Stancel, G.M. & Deal, Jr., W.C. (1969). Reversible dissociation of yeast glyceraldehyde 3-phosphate dehydrogenase by adenosine triphosphate. *Biochemistry* 8, 4005-4011.
- Steinmuller, K., Kaling, M. Zetsch, K. (1983). *In vitro* synthesis of phycobiliproteins and ribulose-1.5-bisphosphate carboxylase by non-polyadenylated-RNA of *Cyanidium caldarium* and *Phorphyridium aeruginum*. *Planta* 159, 308-313.

- Sternberg, N. (1973). Properties of a mutant of *Escherichia coli* defective in bacteriophage 1 head formation (*groE*), II. The propagation of phage 1. *J. Mol. Biol.* 76, 25-44.
- Streusand, V.J. & Portis, A.R.Jr (1987). Rubisco activase mediates ATP-dependent activation of ribulose biphosphate carboxylase. *Plant Physiol.* 85, 152-154.
- Subramanian, A.R., Wittmann-Liebold, B., Geissler, A.W., Stöffler, G. & Giesen, M. (1979). Comparison of ribosomal protein S1 and the A-protein from *Escherichia coli*. *FEBS lett.* 99, 357-360.
- Tabita, F.R. & McFadden, B.A. (1974). D-ribulose-1.5-bisphosphate carboxylase from *Rhodospirillum rubrum*. II Quaternary structure, composition, catalytic and immunological properties. *J. Biol. Chem.* 249, 3459-3464.
- Tabita, F.R. & Small, C.L. (1985). Expression and assembly of active cyanobacterial ribulose-1.5-bisphosphate carboxylase/oxygenase in *Escherichia coli* containing stoichiometric amounts of large and small subunit. *Proc. Natl. Acad. Sci. USA* 82, 6100-6103.
- Tabita, F.R., Gibson, J.L., Mandy, W.J. & Quivey, Jr., R.G. (1986). Synthesis and assembly of a novel recombinant ribulose biphosphate carboxylase/oxygenase. *Biotechnology* 4, 138-141.
- Takano, T. & Kakefuda, T. (1972). Involvement of a bacterial factor in morphogenesis of bacteriophage capsid. *Nature new Biol.* 239, 34-37.
- Terachi, T., Ogihara, Y. & Tsunewaki, K. (1987). Molecular basis of genetic diversity among cytoplasm of Triticum and Aegilops: VI. Complete nucleotide sequences of the *rbl* genes encoding H- and L-type Rubisco large subunit in common wheat and *Ae. crassa* 4x. *Japanese J. Genet.* 62, 375-387.
- Thole, J.E.R., Hindersson, P., de Bruyn, J., Cremers, F., van der Zee, J., de Cock, H., Tommassen, J. van Eden, W. & van Embden, J.D.A. (1988). Antigenic relatedness of a strongly immunogenic 65 kDa mycobacterial protein antigen with a similarly sized ubiquitous bacterial common antigen. *Microbial Pathogenesis* 4, 71-83.
- Thornton, J.M., Edwards, M.S., Taylor, W.R. & Barlow, D.J. (1986). Location of 'continuous' antigenic determinants in the protruding regions of proteins. *EMBO J.* 5, 409-413.
- Tilly, K. & Georgopoulos, C.P. (1982). Evidence that the two *Escherichia coli groE* morphogenetic gene products interact in vivo. *J. Bact.* 149, 1082-1088.

- Tilly, K., Murialdo, H. & Georgopoulos, C.P. (1981). Identification of a second *Escherichia coli* *groE* gene whose product is necessary for bacteriophage morphogenesis. *Proc. Natl. Acad. Sci. USA* 78, 1629-1633.
- Timko, M.P., Kausch, A.P., Castersana, C., Fassler, J., Herrera-Estrella, van den Broeck, G., van Montagu, M., Schell, J. & Cashmore, A.R. (1985). Light regulation of plant gene expression by an upstream enhancer-like element. *Nature* 318, 579-582.
- Tobin, E.M. & Silverthorne, J. (1985). Light regulation of gene expression in higher plants. *Ann. Rev. Plant Physiol.* 36, 569-593.
- Tomizawa, J. & Selzer, G. (1979). Initiation of DNA synthesis in *Escherichia coli*. *Ann. Rev. Biochem.* 48, 999-1034.
- Torres-Ruiz, J.A. & McFadden, B.A. (1988). A homolog of ribulose biphosphate carboxylase/oxygenase-binding protein in *Chromatium vinosum*. *Biochem. Biophys. Acta* 261, 196-204.
- Towbin, H., Staehelin, T. & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA* 76, 4350-4354.
- van der Vies, S.M., Bradley, D. & Gatenby, A.A. (1986). Assembly of cyanobacterial and higher plant ribulose biphosphate carboxylase subunits into functional homologous and heterologous enzyme molecules in *Escherichia coli*. *EMBO J.* 5, 2439-2444.
- Van Dijk, T.K., Gatenby, A.A. & LaRossa, R.A. (1989). Genetic suppression demonstrates interaction of the *groE* products with many proteins (*Nature*, in press).
- van Eden, W., Thole, J.E.R., van der Zee, R., Noordzij, A., van Embden, J.D.A., Hensen, E.J. & Cohen, I.R. (1988). Cloning of the mycobacterial eptope recognised by T-cells in adjuvant arthritis. *Nature* 331, 171-173.
- Viale, A.M., Kobayashi, H., Takabe, T. & Akazawa, T. (1985). Expression of genes for subunits of plant-like Rubisco from *Chromatium* and production of the enzymically active molecule in *Escherichia coli*. *FEBS Lett.* 192, 283-288.
- Vieira, J. & Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19, 259-268.

- Vodkin, M.H. & Williams, J.C. (1988). A heat shock operon in *Coxiella burnetii* produces a major antigen homologous to a protein in both mycobacteria and *Escherichia coli*. *J. Bact.* **170**, 1227-1234.
- Vogt, V.M. (1973). Purification and further properties of single-strand specific nuclease from *Aspergillus oryzae*. *Eur. J. Biochem.* **33**, 192-200.
- von Heijne, G.T. (1987). Sequence analysis in molecular biology. Treasure trove or trivial pursuit. Academic Press Inc. (London) Ltd.
- Voordouw, G., van der Vies, S.M., Bouwmeister, P. (1984). Dissociation of ribulose-1.5-bisphosphate carboxylase/oxygenase from spinach by urea. *Eur. J. Biochem.* **141**, 313-318.
- Wada, M. & Itikawa, H. (1984). Participation of *Escherichia coli* K-12 *groE* gene products in the synthesis of cellular DNA and RNA. *J. Bact.* **157**, 694-696.
- Wales, R., Newman, B.J., Pappin, D. & Gray, J.C. (1989). The extrinsic 33 kDa polypeptide of the oxygen-evolving complex of photosystem II is a putative calcium-binding protein and is encoded by a multi-gene family in pea. *Plant Mol. Biol.* **12**, 439-451.
- Walker J.E., Saraste, M.S., Runswick, M.J. & Gray, N.J. (1982). Distantly related sequences in the a- and b-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **8**, 945-951.
- Walker, D.A. (1973). Photosynthetic induction phenomena and the light activation of ribulose diphosphate carboxylase. *New Phytol.* **72**, 209-235.
- Walter, P & Lingappa, V.R. (1986). Mechanism of protein translocation across the endoplasmic reticulum membrane. *Ann. Rev. Cell Biol.* **2**, 499-516.
- Walter, P., Ibrahimi, I. & Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to *in vitro* assembled polysomes synthesizing secretory protein. *J. Cell Biol.* **91**, 545-550.
- Weigle, J. (1966). Assembly of phage I *in vitro*. *Proc. Natl. Acad. Sci.* **55**, 1462-1466.
- Weiss, B. (1976). Endonuclease II of *Escherichia coli* is exonuclease III. *J. Biol. Chem.* **251**, 1896-1901.
- Westof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A. & van Regenmortel, M.H.V. (1984). Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature* **311**, 123-126.

- Vodkin, M.H. & Williams, J.C. (1988). A heat shock operon in *Coxiella burnetii* produces a major antigen homologous to a protein in both mycobacteria and *Escherichia coli*. *J. Bact.* **170**, 1227-1234.
- Vogt, V.M. (1973). Purification and further properties of single-strand specific nuclease from *Aspergillus oryzae*. *Eur. J. Biochem.* **33**, 192-200.
- von Heijne, G.T. (1987). Sequence analysis in molecular biology. Treasure trove or trivial pursuit. Academic Press Inc. (London) Ltd.
- Voordouw, G., van der Vies, S.M., Bouwmeister, P. (1984). Dissociation of ribulose-1.5-bisphosphate carboxylase/oxygenase from spinach by urea. *Eur. J. Biochem.* **141**, 313-318.
- Wada, M. & Itikawa, H. (1984). Participation of *Escherichia coli* K-12 *groE* gene products in the synthesis of cellular DNA and RNA. *J. Bact.* **157**, 694-696.
- Wales, R., Newman, B.J., Pappin, D. & Gray, J.C. (1989). The extrinsic 33 kDa polypeptide of the oxygen-evolving complex of photosystem II is a putative calcium-binding protein and is encoded by a multi-gene family in pea. *Plant Mol. Biol.* **12**, 439-451.
- Walker J.E., Saraste, M.S., Runswick, M.J. & Gray, N.J. (1982). Distantly related sequences in the a- and b-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **8**, 945-951.
- Walker, D.A. (1973). Photosynthetic induction phenomena and the light activation of ribulose diphosphate carboxylase. *New Phytol.* **72**, 209-235.
- Walter, P & Lingappa, V.R. (1986). Mechanism of protein translocation across the endoplasmic reticulum membrane. *Ann. Rev. Cell Biol.* **2**, 499-516.
- Walter, P., Ibrahimi, I. & Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to *in vitro* assembled polysomes synthesizing secretory protein. *J. Cell Biol.* **91**, 545-550.
- Weigle, J. (1966). Assembly of phage I *in vitro*. *Proc. Natl. Acad. Sci.* **55**, 1462-1466.
- Weiss, B. (1976). Endonuclease II of *Escherichia coli* is exonuclease III. *J. Biol. Chem.* **251**, 1896-1901.
- Westof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A. & van Regenmortel, M.H.V. (1984). Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature* **311**, 123-126.

- Young, R.A. & Davis, R.W. (1983b). Yeast RNA polymerase II genes: isolation with antibody probes. *Science* 222, 778-782.
- Zhu, X., Ohta, Y., Jordan, F. & Ionouye, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* 339, 483-484.
- Zhu, Y.S., Lovett, P.S., Williams, D.M. & Kung, S.D. (1984). *Nicotiana* chloroplast genome. 7. Expression in *E.coli* and *B.subtilis* of tobacco and *Chlamydomonas* chloroplast DNA sequences coding for the large subunit of RuBP carboxylase. *Theor. Appl. Genet.* 67, 333-336.
- Zurawski, G., Perrot, R., Bottomley, W. & Whitfield, P.R. (1981). The structure of the gene for the large subunit of ribulose 1.5-bisphosphate carboxylase from spinach chloroplast DNA. *Nucl. Acids Res.* 14, 3251-3270.
- Zweig, M. & Cummings, D.J. (1973). Cleavage of head and tail proteins during bacteriophage T5 assembly: selective host involvement in the cleavage of a tail protein. *J. Mol. Biol.* 80, 505-518.
- Zylicz, M., Ang, D., Liberek, K. & Georgopoulos, C. (1989). Initiation of lambda DNA replication with purified host- and bacteriophage-encoded proteins: the role of dnaK, dnaJ and groE heat shock proteins. *EMBO J.* 8, 1601-1608.

## 6 INDEX



# Homologous plant and bacterial proteins chaperone oligomeric protein assembly

Sean M. Hemmingsen\*, Carol Woolford†, Saskia M. van der Vies‡, Kit Tilly§,  
David T. Dennis¶, Costa P. Georgopoulos§, Roger W. Hendrix\* & R. John Ellis†

\* Plant Biotechnology Institute, National Research Council, Saskatoon, Saskatchewan, S7N 0W9, Canada

† Department of Biological Sciences, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, USA

‡ Department of Biological Sciences, University of Warwick, Coventry, CV4 7AL, UK

§ Laboratory of Biochemistry, NCI-NIH, Bethesda, Maryland 20892, USA

¶ Department of Biology, Queen's University, Kingston, Ontario, K7L3N6, Canada

§ Department of Cellular, Viral, and Molecular Biology, University of Utah Medical Center, Salt Lake City, Utah 84132, USA

\* Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

*An abundant chloroplast protein is implicated in the assembly of the oligomeric enzyme ribulose biphosphate carboxylase-oxygenase, which catalyses photosynthetic CO<sub>2</sub>-fixation in higher plants. The product of the Escherichia coli groEL gene is essential for cell viability and is required for the assembly of bacteriophage capsids. Sequencing of the groEL gene and the complementary cDNA encoding the chloroplast protein has revealed that these proteins are evolutionary homologues which we term 'chaperonins'. Chaperonins comprise a class of molecular chaperones that are found in chloroplasts, mitochondria and prokaryotes. Assisted post-translational assembly of oligomeric protein structures is emerging as a general cellular phenomenon.*

THE process of post-translational assembly of polypeptides into oligomeric structures is an often overlooked aspect of the molecular biology of gene expression. The assembly process is generally assumed to result from properties inherent in the structures of the polypeptide subunits themselves, because several striking examples of this principle of self-assembly are known<sup>1</sup>. In at least some cases, however, structural information resident in other protein molecules is required for the correct assembly of oligomeric proteins. These assembly proteins have been termed 'molecular chaperones' because they prevent the formation of incorrect structures but are not part of the final oligomer<sup>2</sup>.

Correct post-translational assembly of the head proteins of bacteriophages lambda and T4, and of the tail proteins of bacteriophage T5, requires the function of the *Escherichia coli* groEL gene product<sup>3-6</sup>. A protein in the stroma of higher plant chloroplasts, termed the Rubisco subunit binding protein (referred to later simply as 'binding protein'), is implicated in the post-translational assembly of the enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco)<sup>7,8</sup>. We report here that, based on the similarity of their predicted amino-acid sequences, the groEL gene product and the Rubisco subunit binding protein are evolutionary homologues. Data presented here, and additional data in press, indicate that the proteins of this highly conserved class are ubiquitous, or nearly so. The conservation and ubiquity of this class of proteins indicate that its specific roles in the assembly of Rubisco and of bacteriophage structural proteins reflect a more general and essential cellular function. We propose that this class of proteins be termed 'chaperonins'.

## The groE gene products

The *E. coli* groE locus was first identified in studies of mutations that interfere with the assembly of the heads of bacteriophages lambda and T4 (refs 3, 4, 9, 10). It was subsequently shown that the groE products are also essential for bacteriophage T5 tail assembly<sup>6</sup>. The groE locus maps to 94 min on the *E. coli* genetic map<sup>11</sup> and is composed of two genes, groEL and groES<sup>12</sup>, both of which are essential for cell viability (Fayet, O, Ziegler-Hoffer, T., C.P.G., personal communication). The groEL and groES genes encode polypeptides with apparent relative molecular masses (*M<sub>r</sub>*) of 65,000 and 15,000 respectively<sup>12,13</sup>, which are among the most abundant proteins in the cell. Both

the groEL and groES polypeptides are found in the cell as oligomeric structures. The purified groEL protein is a 14 subunit homo-oligomer composed of two stacked rings of 7 subunits each<sup>14,15</sup>. The groES protein consists of 6-8 identical subunits arranged in a single ring<sup>16</sup>. The groEL protein is an ATPase<sup>14</sup>, and the groEL and groES proteins form a complex with each other in the presence of Mg-ATP but not in its absence<sup>16</sup>. The groE proteins act at an early stage in the head assembly pathways of bacteriophages lambda, T4, and others<sup>17</sup>. In the lambda head assembly pathway, both groE proteins are required for the earliest assembly step that has been detected, which is the formation of an oligomer of 12 subunits of phage protein gpB<sup>5</sup>; this oligomer subsequently forms the 'connector' or 'portal' structure on which head-shell assembly occurs. In the absence of groEL function, the T4 major head protein is found associated with the cell membrane in amorphous 'lumps'. Although the groE proteins participate in normal phage head assembly, they are not found in the assembled head structures<sup>18</sup>.

The groE genes are members of the *E. coli* heat shock regulon. Synthesis of groEL protein, which accounts for 1% of cellular protein synthesis in cells in steady state growth at 37 °C, increases to 10% of total synthesis soon after the cells are shifted to a growth temperature of 46 °C. The promoter controlling transcription of the groE genes is recognized by RNA polymerase carrying the heat shock sigma factor,  $\sigma$  32 (ref. 20).

The DNA sequence of the groE operon is presented in Fig. 1. Examination of this sequence shows two open reading frames encoding polypeptides of *M<sub>r</sub>* 10,368 and *M<sub>r</sub>* 57,259 and predicted *pI* values of 5.92 and 5.63 respectively. These estimates correspond with the respective measured values for the groES and groEL polypeptides. Comparison of amino-acid compositions and amino-terminal sequences determined from the purified groES and groEL proteins with the compositions and sequences deduced from the DNA sequence show excellent agreement. Northern blot experiments (not shown) reveal a single transcript of ~2,100 nucleotides containing both groES and groEL sequences. This transcript, which is four to five times more abundant in extracts of heat-shocked cells compared to non-heat-shocked cells, extends from the previously identified promoter upstream of groES to a predicted terminator stem-loop structure ~60 nucleotides 3' from the groEL coding region. The groEL and groES genes are therefore co-transcribed as an operon.

**Fig. 1** Nucleotide sequence of the *E. coli* *groE* operon. The predicted amino-acid sequences of the *groES* and *groEL* open reading frames are indicated, as well as the promoter and transcription start site<sup>20</sup>, and the dyad symmetry of the proposed transcription terminator. The predicted amino-terminal sequences match the chemically determined sequences of the *groES* and *groEL* proteins. (For *groES*, 8 of 8 residues match, including the amino-terminal methionine (data not shown); 21 of 23 residues match for *groEL*<sup>47</sup>. The amino-terminal methionine is absent in mature *groEL* protein.) The source of DNA for sequencing was the 7.7 kilobase (kb) insert in the *groE* transducing phage H18 (ref. 13). Parts of this DNA were subcloned into plasmid pBR322 and the sequence was determined either directly by the method of Maxam and Gilbert<sup>48</sup> (residues 1-956 and 1953-2198) or by the method of Sanger *et al.*<sup>49</sup> following further subcloning into phage M13 vectors (residues 644-2267). The sequences of both strands of the DNA were determined, and all restriction sites were crossed.

The enzyme Rubisco occurs in higher plant chloroplasts as an oligomer of eight large subunits, which are synthesized within the chloroplast, and eight small subunits, which are synthesized by cytosolic ribosomes<sup>7</sup>. The Rubisco subunit binding protein was first identified in studies with isolated intact chloroplasts of *Pisum sativum* (pea) as a protein that binds newly synthesized Rubisco large subunits before their appearance in the holo-enzyme<sup>7</sup>. The binding protein also non-covalently binds newly synthesized Rubisco small subunits after their import from the cytosol<sup>8</sup>. The binding protein is not itself a component of the assembled Rubisco enzyme.

The binding protein consists of two types of subunit of apparent  $M_r$  61,000 (alpha) and  $M_r$  60,000 (beta)<sup>21</sup>. The subunits are distinct by a number of criteria<sup>21</sup>, but amino-terminal amino-acid sequence analysis shows some sequence similarity<sup>22</sup>. The subunits are encoded by nuclear genes and are imported into the chloroplast after synthesis in the precursor form by cytosolic ribosomes<sup>23</sup>. The purified oligomeric binding protein has an apparent  $M_r$  estimated by gel-sieving to be greater than 700,000 (ref. 23). The binding protein oligomer dissociates reversibly *in vitro* in the presence of Mg-ATP and can be recovered in forms ranging from the high  $M_r$  forms to monomers<sup>21,23-25</sup>. The purified protein has been reported to show ATPase activity<sup>26</sup>.

There is evidence that the binding protein plays an essential role in the assembly of Rubisco in higher plants<sup>8</sup>. The most

convincing evidence is the inhibition of the transfer of newly synthesized Rubisco large subunits to the holoenzyme in stromal extracts by the addition of antiserum to the binding protein<sup>27,28</sup>. In contrast to the higher plant Rubisco, purified Rubisco from cyanobacteria can be dissociated into its subunits and subsequently reassembled into an active enzyme *in vitro*<sup>29</sup>. An active enzyme is also produced when the cyanobacterial Rubisco genes are expressed in *E. coli*<sup>30</sup>. Efforts in several laboratories to extend these results to higher plant Rubisco have met with failure however<sup>30</sup>. The large subunit of higher plant Rubisco is apparently incapable of participating in assembly either after dissociation from the enzyme or after heterologous expression in the absence of the binding protein<sup>29,30</sup>. The failure of assembly in *E. coli* is blocking attempts to produce improved forms of this agriculturally important enzyme by *in vitro* mutagenesis<sup>31</sup>.

Clones encoding castor-bean (*Ricinus communis*) and wheat (*Triticum aestivum*) binding proteins were isolated from cDNA libraries constructed in the expression vector  $\lambda$ gt11 and screened with a specific polyclonal antibody raised against the binding protein from *P. sativum* leaf chloroplasts<sup>21</sup>. The sequences of the plant cDNAs and the predicted amino-acid sequences are presented in Fig. 2.

The wheat cDNA sequence contains an open reading frame encoding a polypeptide with a predicted *pI* of 5.63. The predicted amino-terminal sequence corresponds well with that determined for the amino-terminus of the purified mature alpha

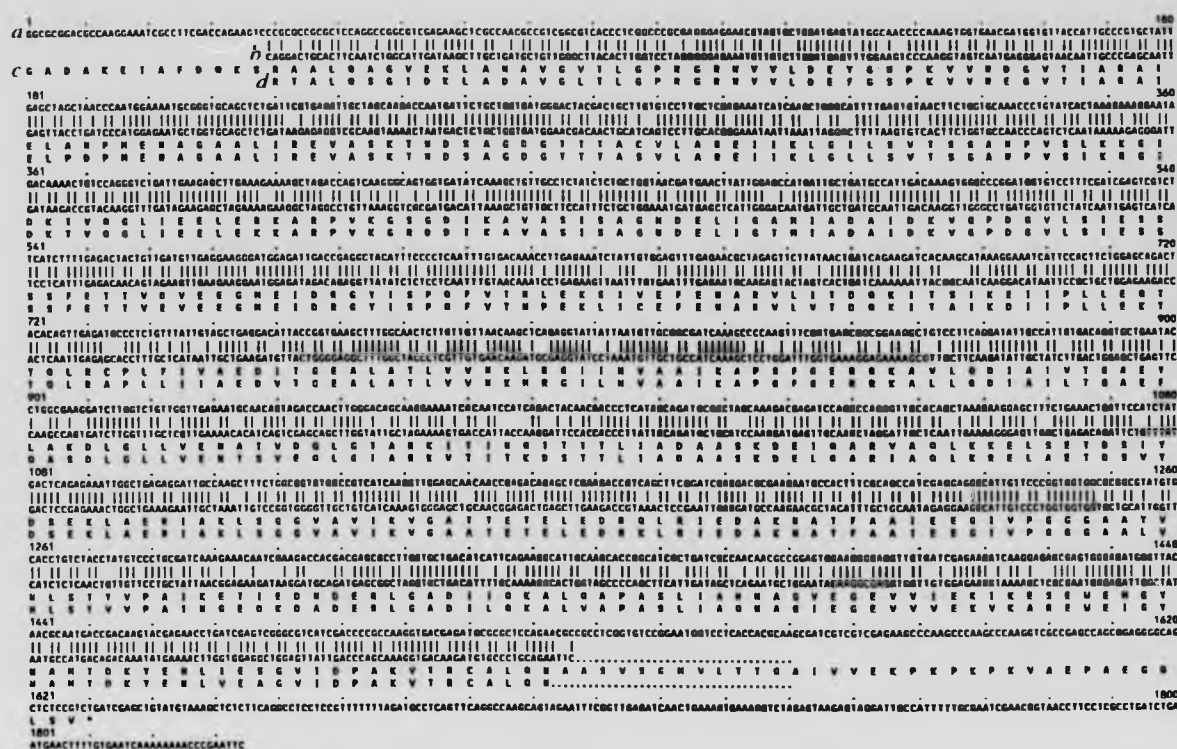


Fig. 2 Nucleotide sequences of the *R. communis* and *T. aestivum* Rubisco subunit binding-protein alpha subunits. The *T. aestivum* (a) and *R. communis* (b) nucleotide sequences and the predicted amino-acid sequences of the open reading frames are indicated (*T. aestivum*, (c) and *R. communis*, (d)). Vertical lines indicate nucleotide identity.

**Methods.** Libraries of cDNA derived from poly(A<sup>+</sup>) RNA extracted from the endosperm of developing *R. communis* seeds and from young wheat leaves, were used to construct  $\lambda$ gt11 expression libraries<sup>40</sup>. The *T. aestivum* library was kindly provided by C. Raines and T. A. Dyer. The *R. communis* library was made without methylation of the cDNA before *Eco*RI digestion and insertion into the vector. Both libraries were screened with a specific polyclonal antibody raised against purified binding protein from *P. sativum* leaf chloroplasts<sup>21</sup>. From 150,000 *R. communis* library recombinants screened, 26 positives were recovered. The inserts from four of these were subcloned into a plasmid vector for further analysis. Each of the subclones hybridized specifically to a transcript of 2,300 bp in total *R. communis* RNA. Sequence determination of the *R. communis* clones was by primer extension using the chain termination method of Sanger *et al.*<sup>40</sup>. A recombinant isolated from the *T. aestivum* expression library contained a cDNA insert of ~2 kb, which hybridizes very strongly to the isolated *R. communis* cDNAs. This fragment was subcloned in the plasmid vector PUBS (a derivative of Bluescript kindly provided to us by G. Murphy) giving rise to pSV10 and pSV11 (opposite orientation); pSV10 and pSV11 were used to determine the *T. aestivum* nucleotide sequence<sup>51</sup>.

polypeptide of the binding protein from both *T. aestivum* (17 matches out of 20) and *P. sativum* (13 matches out of 20). There is less correspondence between the predicted wheat sequence and the amino-terminal sequence of the purified mature beta subunit of *P. sativum* (9 matches out of 20) and four of these matches are seen only if account is taken of two extra residues in the pea beta subunit (Fig. 3). We conclude that the wheat cDNA clone encodes the entire mature binding protein alpha subunit plus two amino acid residues of the presequence. The calculated  $M_r$  of the mature alpha subunit is 57,393. The castor-bean cDNA sequence contains an internal 3' *Eco*RI site and it does not extend to the mature amino-terminus corresponding to that determined for the *P. sativum* alpha subunit; it does, however, include 91% of the coding sequence for the mature polypeptide based on the wheat sequence. The wheat and castor-bean sequences are 80% identical at the amino-acid level.

#### Similarity of plant and bacterial sequences

The predicted amino-acid sequences of the wheat binding-protein alpha subunit and the *groEL* protein are compared in

Fig. 4. The overall amino-acid compositions are very similar and the calculated  $M_r$  and  $pI$  values are almost identical. With the introduction of only a few gaps, 46% of the residues are identical and many of the differences represent conservative substitutions. On the basis of the high sequence similarity, we conclude that the binding and *groEL* proteins are homologues.

An earlier report showed that total leaf extracts of *P. sativum* contain a protein similar to the *groEL* protein in that it consists of 14 subunits, arranged in two layers of 7 monomers each, and shows weak ATPase activity<sup>31</sup>. It seems likely that the protein these authors observed is identical with the Rubisco subunit binding protein. We suggest that the binding protein consists of seven alpha and seven beta subunits, and this extends the similarity between the *groEL* protein and the binding protein to the level of ultrastructure.

The plant and bacterial proteins both show weak ATPase activities and the oligomeric structures of both are affected by the presence of Mg-ATP. The oligomeric structures may be stabilized by hydrophobic interactions upon which the Mg-ATP acts. The dissociation of the plant protein is enhanced at lower

a DAKEIAFDQK SRAALQAGVE  
 \* \* \* \* \*  
 b XAPEIAFDQK SRAALQAGVE  
 \* \* \* \* \*  
 c AAKDIAFDQH SRSAMQAGID  
 \* \* \* \* \*  
 d AREIAFDQKS RAA..LQAGV  
 \* \* \* \* \*  
 e AKELHFNKDG SAIRKLQNGV

Fig. 3 Comparison of the amino-terminal sequences of the binding-protein subunits from *T. aestivum* and *P. sativum*. A lyophilized sample of the *T. aestivum* binding-protein alpha subunit (kindly provided by A. J. Keys) was subjected to automated solid-phase Edman degradation by the SERC protein sequencing unit at the Department of Biochemistry, University of Leeds. Symbols: a, predicted sequence of cDNA from *T. aestivum*; b, determined sequence of binding-protein alpha subunit from *T. aestivum*; c, determined sequence of binding-protein alpha subunit from *P. sativum*<sup>21</sup>; d, determined sequence of binding-protein beta subunit from *P. sativum*<sup>21</sup>. Asterisks indicate identical residues, dots indicate gaps inserted to maximize the identity. The three discrepancies between the determined sequence and the predicted sequence of the *T. aestivum* subunit (lines a and b) may result from incomplete recovery of aminoterminal and lysine residues associated with the sequencing method; alternatively, they may reflect varietal differences, because the alpha subunit was obtained from variety 'Ficco' while the cDNA was prepared from variety 'Chinese Spring'.

temperature<sup>23</sup> and the bacterial protein will dissociate in metrizamide, which is less polar than water, in the presence of Mg-ATP (Chandrasekhar, G. N. & C.P.G., personal communication). The requirement for Mg-ATP for the association of the *groEL* and *groES* proteins may in fact be a requirement for the destabilisation of the interactions of *groEL* monomers amongst themselves. These similarities suggest that the proteins retain similar physical and functional properties throughout evolution.

### Ubiquitous occurrence

Bacterial DNA sequences with a high degree of similarity to the *groEL* coding sequence have been identified in a number of other bacterial species<sup>32</sup>. The *groEL*-like protein is a major immunogen in a number of infections including those caused by mycobacteria. Studies using antibodies to the antigen from *Mycobacterium leprae* and *M. tuberculosis* have shown that this antigen is extremely widespread in prokaryotes, including archaeobacteria<sup>33,34</sup>. Antibodies to the *P. sativum* binding protein have been used independently and recognize a polypeptide of apparent *M*, 60,000 in similarly diverse groups of bacteria (S.M.H., S.M.V. & R.J.E., unpublished observations).

An abundant protein with an apparent subunit *M*, of 58,000 has been purified recently from *Tetrahymena thermophila* and shown to be related to the *groEL* protein both immunologically and structurally<sup>35,36</sup>. This *groEL*-like protein is localized in the mitochondria, where its suggested function is to assist in the refolding of newly imported proteins. Antibodies to the *Tetrahymena groEL*-like protein cross-react with a protein of apparent subunit *M*, of 58-64,000 in extracts of mitochondria from yeast, *Xenopus*, maize and human, suggesting that it occurs in all mitochondria<sup>36</sup>.

Of the three bacterial strains with sequenced *groEL* homologues, one, *Coxiella burnetii*, has a homologue of the *groES* gene, which, as in the *E. coli* case, lies a short distance upstream from the *groEL* gene (Vodkin, M. & Williams, J., personal communication). The *C. burnetii groES* protein shows 53% amino-acid sequence identity with the *E. coli groES* protein. For the two mycobacterial species, there is no evidence for sequence matches to the *groES* gene, despite the availability of several hundred base pairs (bp) of sequence on both sides of the *groEL* coding regions. There is no information on the possible occurrence of the *groES* sequence in eukaryotes.

a 1 GADAKEIAFDQK SRAALQAGVE KLANAVGVTLCGRNVLDE. YGNFKV 49  
 b 1 .MAAKDVKFGNDARVKNLRGVNLADAVKVTLPKGRNVLDKSGFAPTTI 49  
 50 VNDGVITIRATEIANPHEAGALIREVASRTNDSAGDTTTCVLAEREI 99  
 50 TKDGVSVARETELEDKFFENMGAGMVEKASKANDAGIXGNNTATVLAQAI 99  
 100 IKLGILSVTSGANPVSLKRGIDKTVOGLIEELERKAPVKGSGUIKAVAS 149  
 100 ITTEGLKAVAAGNPMDLKRGIDKAVTAAVEELKALSVFCSKATAQVGT 149  
 150 ISAGNDELIGAMIDATDKVPGDGLSIESSSFETTVDVEFGMEIDRGV 199  
 150 ISATSEDTVGKLIAEAMDKVGKGVITVEDGTGLQDELDVVEGMDRGV 199  
 200 ISPOFVTHLEKSI VEFENARVLITDQKITSIKEIIPLLIQTOTIRCFIFI 249  
 200 LSPVFINPKPETGAVELESFFILLADKISINIREMLPULAVAKAGKPLLI 249  
 250 VAFDITGEALATLVNKLKRGIIINVAIKAPSYGERRKAVIQDIAIVTGA 299  
 250 IAEDEVEGEALATAVVNTIRGIVKVAVKAFGFGDRKAMLDQIATLTGGT 299  
 300 VLAQDLGLLENATVDQLGTARKITIHOTTTTLADAASKDEIGARVAQL 349  
 300 VISEEIGMELEKATLEDLGAKRVVINKDTTITDGVGEAAIQGRVAQI 349  
 350 KKESETDSIYDSEKIAFRIAKISGGVAVIKVCATTETELDROIRIFDA 399  
 350 RQOIEEATSDVDREKLEQERVAKLGGVAVIKVGAATEVEMKKAVEREDA 399  
 400 KNATFAATIEEGIVPGGAAYVHLSTYVPATKETIEDHDERIGADITQKAL 449  
 400 LHATRAAVEEGVAVAGGVVALIRVASKLADLRG. QNEDQNVGIKVALRAM 447  
 450 OAPAGLIANNAGVEGEVVEIKIESEWEMGYNAMTDKYNLIESGVIDFA 499  
 448 EAPLRQIVLNCGEPSVAVANTVKGDDNGYGVNAATEYGNHIDNGILDTT 497  
 500 KVTRECALQNAASVSGNVLTTQAIIVKPKPKPKVAEPAGQISV\* 544  
 498 KVTREALQVAASVAGIMITTECHVVDLPKNDADLGAAGGNGMGNGMGH\* 549

Fig. 4 Predicted amino-acid sequences of the wheat binding protein alpha subunit (a) and the *E. coli groEL* protein (b). Vertical lines indicate amino-acid identity, and dots indicate gaps inserted to maximize identity. The first two residues of the wheat sequence are the last two residues of the presequence.

### Discussion

We have described a ubiquitous, conserved, abundant protein that is associated with the post-translational assembly of at least two structurally distinct oligomeric protein complexes. We conclude that the role of this protein is to assist other polypeptides to maintain or to assume conformations which permit their correct assembly into oligomeric structures. We propose that this protein be called a 'chaperonin', since its properties meet the criteria suggested for molecular chaperones<sup>2</sup>. The term 'molecular chaperone' was first used to describe nucleoplasmin, an acidic nuclear protein required for the assembly of nucleosome cores from DNA and histones in extracts of *Xenopus* eggs<sup>37</sup>. The term has now been extended to describe a family of proteins whose proposed role is to ensure that the folding of certain polypeptides and their assembly into oligomeric structures occur correctly<sup>2</sup>. It has been suggested that the essential function of molecular chaperones is to prevent the formation of 'improper' structures which might result from the transient exposure of hydrophobic or charged surfaces normally involved in domain interactions, either within or between polypeptide chains. Such exposure may occur during the operation of several cellular processes. These include the synthesis of polypeptides, the refolding that occurs after their transport across membranes, the association of polypeptides made in one subcellular compartment with those made in another, changes in protein-protein interactions during the normal functioning of a complex and recovery from stress such as heat shock. The other criterion that distinguishes molecular chaperones is that they do not form part of the final structure whose assembly they have mediated. We recognize currently three classes within the molecular chaperone family, namely, (1) nucleoplasmin, (2) *hsp70*-immunoglobulin heavy chain binding protein class, and (3) the bacterial-mitochondrial-chloroplast class. We propose to limit the term 'chaperonin' to the last class.

Some of the chaperonins in both prokaryotes and eukaryotes are known to be heat-shock proteins<sup>19,35,36</sup>. It has been proposed recently that heat-shock proteins may function both to refold heat-denatured proteins and to influence the folding or aggregation state of other proteins under non-stress conditions<sup>38</sup>. This proposal supports the functions we have suggested for the molecular chaperones.

Some of the proteins most highly conserved between prokaryotes and eukaryotes are found associated with mitochondria and chloroplasts, for example the beta subunit of ATP synthetase<sup>39</sup> and elongation factor EFTu<sup>40</sup>. As with these proteins, the occurrence of the highly conserved chaperonins in both mitochondria and chloroplasts, as well as in prokaryotes, is most easily explained in terms of the endosymbiotic hypothesis for the origin of these organelles. We suggest that the original role of prokaryotic chaperonin has expanded during evolution to include the mediation of processes peculiar to the biogenesis of these organelles, especially the refolding of polypeptides imported from the cytosol and their combination with polypeptides synthesized within the organelle. We note that the high level of sequence identity that we report here between the prokaryotic and eukaryotic chaperonins is very similar to the levels of identity reported elsewhere between the prokaryotic and eukaryotic members of the *hsp70* and *hsp83* heat-shock protein classes<sup>41,42</sup>. It thus seems possible that the eukaryotic heat-shock proteins as a class were derived from prokaryotes endosymbiotically, even though the *hsp70* and *hsp83* proteins are not now located in organelles.

It remains to consider what the normal roles of chaperonin might be in prokaryotes in the absence of phage infection or environmental stresses. Deletion mutations of the *groE* operon cannot be isolated at temperatures ranging from 20 °C to 43 °C, demonstrating that the *groE* proteins are essential for *E. coli* growth at all temperatures (Fayet, O., Ziegelhoffer, T., C.P.G., personal communication). Mutations in the *groEL* and *groES* genes interfere with DNA and RNA synthesis at high temperatures<sup>43</sup>, while overproduction of both the *groEL* and *groES* proteins from a plasmid suppresses certain mutations in the *dnaA* gene, which is required for the initiation of DNA replication<sup>44,45</sup>. Similarly, the presence of a plasmid carrying what now appears to be the *groE* operon (or even one carrying an internal fragment of the *groEL* gene) suppresses a temperature-sensitive mutation in *ams*, an *E. coli* gene that influences messenger RNA half-life<sup>46</sup>. One interpretation of these observations is that

the *groEL* and *groES* chaperonins function normally to mediate protein-protein interactions involved in nucleic acid metabolism in *E. coli*, a similar suggestion has been made to explain the sequence similarity between the major heat shock proteins of animal cells (*hsp70*) and the product of the *dnaK* gene in *E. coli*<sup>38</sup>. It is also conceivable that the chloroplast and mitochondrial chaperonins function in DNA replication within these organelles and that the chaperonins mediate the assembly of oligomeric complexes other than those involved in nucleic-acid metabolism. Future work will aim to define the scope of chaperonin action in the formation of protein complexes in both prokaryotes and eukaryotes. The idea that molecular chaperones are involved transiently in the folding or assembly of certain oligomeric proteins also has implications for those biotechnologists who are encountering problems in producing foreign proteins in bacterial cells in the required active form. It may be necessary to re-examine the natural synthesis of these proteins by techniques which permit the detection of noncovalently bound complexes to determine whether molecular chaperones are involved.

The involvement of the chloroplast chaperonin in the assembly of the higher plant Rubisco could account for the failure to obtain the synthesis of enzymatically active Rubisco oligomers in *E. coli*<sup>40</sup>. It would follow that the bacterial chaperonin must be sufficiently different from the chloroplast chaperonin to be unable to mediate the assembly of the Rubisco from higher plants. The observation that Rubisco from cyanobacteria can be synthesized in active oligomeric form by *E. coli*<sup>40</sup> suggests either that the *E. coli* chaperonin is able to mediate the assembly of the Rubisco from prokaryotes or that the assembly of Rubisco in cyanobacteria does not involve chaperonin. Our current aim is to express the cDNAs for the chloroplast chaperonin in the same *E. coli* cells that are expressing the genes for Rubisco from higher plants, in the hope of rescuing Rubisco assembly and so permitting attempts to improve the properties of this agriculturally important enzyme.

We thank D. Bradley, V. Knowles, M. A. McCrae and G. Murphy for assistance with DNA sequencing, J. Keen for determination of the amino-terminal protein sequence, C. Raines and T. A. Dyer for access to a wheat cDNA library, H. F. Kung, D. B. Young and R. A. Young for helpful discussion and SERC and NIH for financial support. The sequence data in this publication are being submitted to the EMBL/GenBank Libraries.

Received 14 March; accepted 20 April 1988.

- Alberts, B. *et al.* *Molecular Biology of the Cell* (Garland, New York, 1983).
- Ellis, R. J. *Nature* **328**, 378-379 (1987).
- Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R. & Kaiser, A. D. *J. molec. Biol.* **76**, 45-60 (1973).
- Sternberg, N. *J. molec. Biol.* **76**, 25-44 (1973).
- Kochan, J. & Muraldo, H. *Virology* **131**, 100-115 (1983).
- Zweig, M. & Cummings, D. *J. molec. Biol.* **80**, 505-518 (1973).
- Barracough, R. & Ellis, R. J. *Biochim. Biophys. Acta* **608**, 19-31 (1980).
- Ellis, R. J. & van der Vies, S. M. *Photosynthesis Res.* (in the press).
- Takano, T. & Kakefuda, T. *Nature new Biol.* **239**, 34-37 (1972).
- Coppo, A., Manzi, A., Pulitzer, J. F. & Takahashi, H. *J. molec. Biol.* **76**, 61-87 (1973).
- Georgopoulos, C. P. & Eisen, H. *J. supramolec. Struct.* **2**, 349-359 (1974).
- Tilly, K., Muraldo, H. & Georgopoulos, C. P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1629-1633 (1981).
- Georgopoulos, C. P. & Hohn, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 131-135 (1978).
- Hendrix, R. W. *J. molec. Biol.* **129**, 375-392 (1979).
- Hohn, T., Hohn, B., Engel, A., Wurtz, M. & Smith, P. R. *J. molec. Biol.* **129**, 359-373 (1979).
- Chandrasekar, G. N., Tilly, K., Woolford, C., Hendrix, R. & Georgopoulos, C. *J. biol. Chem.* **261**, 12414-12419 (1986).
- Georgopoulos, C. P., Tilly, K. & Casjens, S. R. in *Lambda II* (eds Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A.) 279-304 (Cold Spring Harbor Monograph 13, Cold Spring Harbor, 1983).
- Fredman, D. I. *et al.* *Microbiol. Rev.* **48**, 299-325 (1984).
- Neidhardt, F. C. & VanBogelen, R. A. in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*) 1334-1345 (American Society for Microbiology, 1987).
- Cowing, D. W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 2679-2683 (1985).
- Musgrove, J. E., Johnson, R. A. & Ellis, R. J. *Eur. J. Biochem.* **163**, 529-534 (1987).
- Ellis, R. J., van der Vies, S. M. & Hemmingsen, S. M. in *Plant Molecular Biology 1987* (eds von Wettstein, D. & Chua, N.-H.) 33-40 (NATO-ASI Series A, Life Sciences Vol. 140, Plenum, New York and London, 1987).
- Hemmingsen, S. M. & Ellis, R. J. *Pl. Physiol.* **80**, 269-276 (1986).
- Musgrove, J. E. & Ellis, R. J. *Phil. Trans. R. Soc. B313*, 419-428 (1986).
- Bloom, M. V., Milos, P. & Roy, H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1013-1017 (1983).
- Chaudhuri, P., Cannon, S., Hubbs, A. & Roy, H. Abstracts 16th Annual UCLA Symposium, *J. cell Biochem. suppl.* **11B**, 50 (1987).
- Milos, P. & Roy, H. *J. cell Biochem.* **24**, 153-162 (1984).
- Cannon, S., Wang, P. & Roy, H. *J. Cell Biol.* **103**, 1327-1335 (1986).
- Andrews, T. J. & Lottner, G. H. *J. biol. Chem.* **260**, 4632-4636 (1985).
- Bradley, D., van der Vies, S. M. & Gatenby, A. A. *Phil. Trans. R. Soc. B313*, 447-458 (1986).
- Pushkin, A. V. *et al.* *Biochim. Biophys. Acta* **704**, 379-384 (1982).
- Young, D., Lathigra, R., Hendrix, R., Sweetser, D. & Young, R. A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Thole, J. E. R. *Microbiol. Pathogenesis* (in the press).
- Young, D. B., Ivanyi, J., Cox, J. H. & Lamb, J. R. *Immunology Today* **8**, 215-219 (1987).
- McMullin, T. W. & Hallberg, R. L. *Molec. cell Biol.* **7**, 4414-4423 (1987).
- McMullin, T. W. & Hallberg, R. L. *Molec. cell Biol.* **8**, 371-380 (1988).
- Laskey, R. A., Honda, B. M., Mills, A. P. & Finch, J. T. *Nature* **275**, 416-420 (1978).
- Pelham, H. R. B. *Cell* **46**, 959-961 (1986).
- Zurawski, G. & Clegg, M. T. *Nucleic Acids Res.* **12**, 2549-2559 (1984).
- Nagata, S., Tsunetsugu-Yokota, Y., Naito, A. & Kaziro, Y. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6192-6196 (1983).
- Bardwell, J. & Craig, E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 848-852 (1984).
- Bardwell, J. & Craig, E. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5177-5181 (1987).
- Wada, M. & Itikawa, H. *J. Biol. Chem.* **263**, 694-696 (1988).
- Fayet, O., Louarn, J.-M. & Georgopoulos, C. *Molec. gen. Genet.* **202**, 435-445 (1986).
- Jenkins, A. J., March, J. B., Oliver, I. R. & Masters, M. *Molec. gen. Genet.* **202**, 446-454 (1986).
- Chanda, P. K., Ono, M., Kuwano, M. & Kung, H.-F. *J. Biol. Chem.* **260**, 446-449 (1985).
- Subramanian, A. R., Wittman-Liebold, B., Geissler, A. W., Stoffler, G. & Giesen, M. *FEBS Lett.* **99**, 357-360 (1979).
- Maxam, A. M. & Gilbert, W. *Methods Enzym.* **65**, 449-560 (1980).
- Sanger, F., Nicklen, S. & Coulson, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5461-5467 (1977).
- Young, R. A. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1194-1198 (1983).
- Murphy, G. & Kavanagh, T. *Nucleic Acids Res.* (in the press).

**THIS VOLUME HAS A**

**VERY TIGHT BINDING**

